

The Dissertation Committee for Stephen Charles Palmer  
certifies that this is the approved version of the following dissertation:

**What has been the preliminary impact of federal  
programs to promote the adoption of electronic health  
records?**

Committee:

---

David Warner, Supervisor

---

Carolyn Heinrich

---

Cynthia Osborne

---

William Sage

---

Sam Richardson

What has been the preliminary impact of federal  
programs to promote the adoption of electronic health  
records?

by

Stephen Charles Palmer, B.A., M.P.Aff

**DISSERTATION**

Presented to the Faculty of the Graduate School of  
The University of Texas at Austin  
in Partial Fulfillment  
of the Requirements  
for the Degree of

**DOCTOR OF PHILOSOPHY**

THE UNIVERSITY OF TEXAS AT AUSTIN

August 2015

# **What has been the preliminary impact of federal programs to promote the adoption of electronic health records?**

Publication No. \_\_\_\_\_

Stephen Charles Palmer, Ph.D.  
The University of Texas at Austin, 2015

Supervisor: David Warner

The objective of this study is to assess the preliminary impact of federal programs established to promote the adoption of EHRs on the patterns, rates, and levels of EHR adoption. Statewide surveys of physicians and hospitals in Texas fielded between 2010 and 2012 are supplemented with other data, and then analyzed with logistic regression modeling, technology diffusion models, and regression discontinuity designs, supplemented with key informant interviews. The models did not consistently yield large or significant effect estimates. The logit models did not show any significant changes in the patterns of EHR adoption. The technology diffusion models showed a small, significant effect on adoption rates. The regression discontinuity (RD) design models showed a moderate, insignificant effect of the programs on adoption levels. Overall, the data and methods used in this study do not support the claim that the federal EHR promotion programs had a significant effect.

# Table of Contents

<b>Abstract</b>	<b>iii</b>
<b>List of Tables</b>	<b>x</b>
<b>List of Figures</b>	<b>xiii</b>
<b>Chapter 1. Introduction</b>	<b>1</b>
1.1 Questions and Hypotheses . . . . .	2
1.2 Policy Relevance . . . . .	6
1.3 Background . . . . .	6
1.3.1 Description of HIT as a Supporting Technology for Clinical Practice and Health Services . . . . .	7
1.3.2 Electronic Health Records: What they are, how they are used, and why it matters . . . . .	7
1.3.3 Appearance of HIT on the Policy Agenda . . . . .	8
1.3.4 Emergence of HIT as a Meaningful Policy Alternative . . . . .	9
1.3.5 Current State of Federal HIT Activity . . . . .	10
1.3.6 EHR Adoption Programs . . . . .	11
1.4 Contribution to the Literature . . . . .	16
1.4.1 Primary contributions to the literature . . . . .	16
1.4.2 Secondary contributions to the literature . . . . .	18
<b>Chapter 2. Literature Review</b>	<b>19</b>
2.1 Physician EHR Adoption . . . . .	19
2.2 Hospital EHR Adoption . . . . .	21
2.3 Determinants of EHR Adoption . . . . .	22
2.4 Impacts of EHR Adoption . . . . .	23
2.5 Diffusion of Innovations . . . . .	24

<b>Chapter 3. Data</b>	<b>26</b>
3.1 Surveys . . . . .	26
3.2 Data Representativeness . . . . .	31
3.3 Key Informant Interviews . . . . .	35
<b>Chapter 4. Methodology</b>	<b>36</b>
4.1 Data Set Development . . . . .	36
4.2 Quantitative Models . . . . .	39
4.2.1 Logistic Regression Models . . . . .	39
4.2.2 Bass Diffusion Model . . . . .	39
4.2.3 Regression Discontinuity Design . . . . .	40
4.2.4 Statistical Software . . . . .	41
<b>Chapter 5. Have patterns in EHR adoption been changing in ways consistent with the programmatic features of the federal EHR promotion programs?</b>	<b>42</b>
5.1 Overview . . . . .	42
5.2 Key Questions . . . . .	44
5.3 Hypotheses . . . . .	44
5.4 Model Development . . . . .	45
5.5 Physician EHR Adoption Patterns . . . . .	46
5.5.1 Data . . . . .	46
5.5.1.1 Survey . . . . .	47
5.5.1.2 Data Set Construction . . . . .	47
5.5.1.3 Key Variables . . . . .	48
5.5.2 Results . . . . .	51
5.5.2.1 Bivariate Analysis . . . . .	51
5.5.2.2 Logistic Analysis . . . . .	56
5.5.3 Hypothesis Test . . . . .	63
5.5.4 Discussion . . . . .	65
5.5.5 Limitations . . . . .	65
5.5.6 Conclusion . . . . .	66
5.6 Hospital EHR Adoption Patterns . . . . .	67

5.6.1	Data . . . . .	67
5.6.1.1	Survey . . . . .	67
5.6.1.2	Data Set Construction . . . . .	68
5.6.1.3	Key Variables . . . . .	68
5.6.2	Results . . . . .	70
5.6.2.1	Bivariate Analysis . . . . .	70
5.6.2.2	Logistic Analysis . . . . .	73
5.6.3	Hypothesis Test . . . . .	79
5.6.4	Discussion . . . . .	80
5.6.5	Limitations . . . . .	81
5.6.6	Conclusion . . . . .	81

**Chapter 6. Have federal programs to promote the adoption of electronic health records changed the rate of adoption? 82**

6.1	Background . . . . .	82
6.1.1	Question . . . . .	84
6.1.2	Method . . . . .	84
6.1.3	Hypothesis . . . . .	84
6.1.4	Theory . . . . .	84
6.1.5	Alternate hypothesis/theory . . . . .	85
6.2	Diffusion of Innovations . . . . .	85
6.3	Approach . . . . .	87
6.4	Past Projections of EHR Adoption Using the Bass Model . . . . .	88
6.4.1	Theoretical Problems with Past Approaches . . . . .	90
6.4.2	Statistical Problems with Past Studies . . . . .	90
6.4.3	Re-creation of Previous Studies . . . . .	92
6.5	Sensitivity Analysis . . . . .	93
6.5.1	Extrema Permutation Analysis . . . . .	94
6.5.2	Monte Carlo Modeling to Test Extrema Permutation Assumptions . . . . .	96
6.6	The Bass Model: a Reprise . . . . .	100
6.7	Building a Better Confidence Interval . . . . .	104

6.8	Hypothesis Test . . . . .	107
6.9	Discussion . . . . .	108
6.10	Limitations . . . . .	109
6.11	Conclusion . . . . .	110
<b>Chapter 7.</b>	<b>Have federal programs to promote the adoption of electronic health records changed the level of adoption?</b>	<b>111</b>
7.1	Overview . . . . .	111
7.2	Regression Discontinuity Design . . . . .	112
7.3	Methodological Note Regarding Linear Models . . . . .	114
7.4	Impact of Regional Extension Centers on Physician EHR Adoption	114
7.4.1	Data . . . . .	115
7.4.2	Theory . . . . .	117
7.4.3	Hypothesis . . . . .	117
7.4.4	Results . . . . .	118
7.4.4.1	Assessing Parametric Form . . . . .	118
7.4.4.2	Assessing the Possibility of Sorting near the Cutoff	120
7.4.4.3	Optimal Bandwidth Calculation . . . . .	122
7.4.4.4	Graphing Different Parametric Forms . . . . .	122
7.4.5	Discussion . . . . .	126
7.4.6	Limitations . . . . .	126
7.5	Hospital Medicaid EHR Incentive Program Impact . . . . .	127
7.5.1	Data . . . . .	128
7.5.2	Theory . . . . .	128
7.5.3	Hypothesis . . . . .	128
7.5.4	Results . . . . .	129
7.5.4.1	Assessing Parametric Form . . . . .	129
7.5.4.2	Assessing the Possibility of Sorting near the Cutoff	132
7.5.4.3	Optimal Bandwidth Calculation . . . . .	134
7.5.4.4	Graphing Different Parametric Forms . . . . .	134
7.5.5	Discussion . . . . .	138
7.5.6	Limitations . . . . .	138
7.6	Conclusion . . . . .	139

<b>Chapter 8. Exploring the attitudes and motivations of health care providers regarding electronic health records and federal programs to promote their adoption</b>	<b>140</b>
8.1 Overview . . . . .	140
8.2 Data . . . . .	141
8.3 Physician Interviews . . . . .	141
8.3.1 Question Variations . . . . .	142
8.3.2 Perceived Impact of EHR Adoption on Quality of Care .	143
8.3.3 Perceived Impact of EHR Adoption on Patient Interactions	143
8.3.4 Perceived Financial Characteristics of EHR Adoption . .	144
8.4 Physician Survey . . . . .	144
8.5 Hospital Survey . . . . .	150
8.6 Discussion . . . . .	153
8.7 Limitations . . . . .	153
8.8 Conclusion . . . . .	154
<b>Chapter 9. Discussion</b>	<b>155</b>
9.1 Impact of Programs on Patterns of Adoption – Logit Analysis	156
9.2 Impact of Programs on Rates of Adoption – Bass Diffusion Model	157
9.3 Impact of Programs on Adoption Levels – RD Design . . . . .	159
9.4 Limitations . . . . .	160
9.5 Review of Key Findings . . . . .	162
9.6 Policy Implications . . . . .	163
9.7 Market Analysis . . . . .	165
9.8 Extensions . . . . .	167
<b>Chapter 10. Conclusion</b>	<b>170</b>
<b>Appendices</b>	<b>173</b>
<b>Appendix A. Physician Logit Analysis Including Hospital Referral Regions</b>	<b>174</b>
<b>Appendix B. Physician EHR Adoption Estimated Using Linear Probability Models</b>	<b>179</b>



Appendix C. Hospital Logit Analysis Including Hospital Referral Regions	182
Appendix D. Hospital EHR Adoption Estimated Using Linear Probability Models	187
Appendix E. Predicting Physician EHR Adoption Using Survival Analysis	190
E.1 Survival Analysis . . . . .	190
E.1.1 Application of Survival Analysis . . . . .	191
E.1.2 Non-Parametric Survival/Hazard Functions . . . . .	191
E.1.3 Proportional Hazard and Parametric Survival Functions	193
E.1.4 Limitations of Survival Analysis for Projection . . . . .	193
Appendix F. Bass Model Sensitivity Analysis by Effect Year	194
Appendix G. Robustness of Physician RD Design Models to Addition of Covariates	198
Appendix H. Robustness of Physician RD Design Models to Different Bandwidth Calculations	202
Appendix I. Robustness of Hospital RD Design Models to Addition of Covariates	203
Appendix J. Robustness of Hospital RD Design Models to Different Bandwidth Calculations	207
Appendix K. R Code Used for Statistics, Tables, and Figures	208
Bibliography	347

## List of Tables

2.1	EHR Adoption Level . . . . .	21
3.1	Primary Data Sources . . . . .	29
3.2	Supplemental Data Sources . . . . .	30
4.1	Hospital Data Set . . . . .	37
4.2	Physician Data Set . . . . .	38
5.1	Summary Table . . . . .	50
5.2	Cross-tabs by EHR Adoption . . . . .	54
5.3	Cross-tabs by Medicaid Participation . . . . .	55
5.4	Physicians, Both Years, Continuous Models . . . . .	57
5.5	Additional Summary Statistics, Continuous Models . . . . .	58
5.6	Physicians, Both Years, Factor Models, Explanatory Variables . . . . .	59
5.7	Physicians, Both Years, Factor Models (cont.), Year Effects . . . . .	60
5.8	Additional Summary Statistics, Factor Models . . . . .	61
5.9	Hospital Data, Summary Statistics, Continuous Variables . . . . .	68
5.10	Hospital Data, Summary Statistics, Nominal Variables . . . . .	69
5.11	Hospital Data, Cross-tabs by EHR Adoption . . . . .	73
5.12	Hospitals, Both Years, Continuous Models . . . . .	75
5.13	”Goodness of Fit” Statistics, Continuous Models . . . . .	76
5.14	Hospitals, Both Years, Factor Models, Explanatory Variables . . . . .	76
5.15	Hospitals, Both Years, Factor Models (cont), Year Effects . . . . .	77
5.16	”Goodness of Fit” Statistics - Factor Models . . . . .	78
6.1	Comparative Residuals . . . . .	92
6.2	Estimates from Ford et al (2006) . . . . .	92
6.3	Estimates from Palmer (2014) . . . . .	92
6.4	Adoption Levels and Standard Errors/Deviations . . . . .	95

6.5	Extrema Permutations . . . . .	95
6.6	Estimated Parameter Values . . . . .	104
6.7	Actual Adoption Levels and Confidence Intervals . . . . .	108
7.1	REC RDD Effect Sizes at Cut-off . . . . .	125
7.2	Medicaid RDD Effect Sizes at Cut-off . . . . .	137
A.1	Physicians, Both Years, Factor Models with HRRs, Table 1 . . .	175
A.2	Physicians, Both Years, Factor Models with HRRs, Table 2 . . .	176
A.3	Physicians, Both Years, Factor Models with HRRs, Table 3 . . .	177
A.4	Physicians, Both Years, Factor Models with HRRs, Table 4 . . .	178
B.1	Physicians, Both Years, Linear Probability Factor Models, Explanatory Variables . . . . .	180
B.2	Physicians, Both Years, Linear Probability Factor Models (cont.), Year Effects . . . . .	181
C.1	Hospitals, Both Years, Factor Models with HRR, Table 1 . . . .	183
C.2	Hospitals, Both Years, Factor Models with HRR, Table 1 . . . .	184
C.3	Hospitals, Both Years, Factor Models with HRR, Table 1 . . . .	185
C.4	Hospitals, Both Years, Factor Models with HRR, Table 1 . . . .	186
D.1	Hospitals, Both Years, LPM Factor Models, Explanatory Variables . . . . .	188
D.2	Hospitals, Both Years, LPM Factor Models (cont), Year Effects . . . . .	189
F.1	Actual Adoption Levels and Confidence Intervals, 2008 Implementation . . . . .	194
F.2	Actual Adoption Levels and Confidence Intervals, 2010 Implementation . . . . .	195
F.3	Actual Adoption Levels and Confidence Intervals, 2011 Implementation . . . . .	195
G.1	REC RDD Effect Sizes at Cut-off with Covariates . . . . .	201
H.1	Expanded REC RDD Effect Sizes at Cut-off . . . . .	202
I.1	Medicaid RDD Effect Sizes at Cut-off with Covariates . . . . .	206

J.1 Medicaid RDD Effect Sizes at Cut-off . . . . .	207
--	-----

## List of Figures

3.1	Comparison of Physician Practice Size Distribution . . . . .	32
3.2	Comparison of Hospital Size Distribution . . . . .	34
5.1	EHR Adoption Level by Year . . . . .	49
5.2	Physician Variable Correlogram . . . . .	52
5.3	Hospital Variables Correlogram, 2012 . . . . .	72
6.1	Adoption of Household Products in the U.S. . . . .	83
6.2	Ford (2006) - Best estimate, optimistic, and conservative . . .	89
6.3	Monte Carlo Results for Minimum p . . . . .	98
6.4	Monte Carlo Results for Maximum q . . . . .	99
6.5	Adopting Physicians by Year . . . . .	102
6.6	Cumulative Distribution . . . . .	103
6.7	Basic Confidence Interval Plots with Bootstrapping . . . . .	106
7.1	Adoption vs Practice Size - Data, Fitted Curve, and Cutoff . .	119
7.2	Practice Size Distribution, Fitted Curve, and Cutoff . . . . .	121
7.3	REC RD Design Table - Multiple Models . . . . .	123
7.4	REC RD design Table - Multiple Models, Cut-off zoom . . . .	124
7.5	Hospital Medicaid % and EHR Adoption - Data, Fitted Curve, and Cutoff . . . . .	130
7.6	Hospital Medicaid % - Histogram, Fitted Curve, and Cutoff .	133
7.7	MCD RD design Table - Multiple Models . . . . .	135
7.8	MCD RD design Table - Multiple Models, Zoom . . . . .	136
8.1	Perceptions regarding EHR Impacts on Quality by Physicians who have Adopted EHRs . . . . .	145
8.2	Perceptions regarding EHR Usability by Physicians who have Adopted EHRs . . . . .	146
8.3	Reasons for Not Adopting as Reported by Non-adopters . . .	147

8.4	Actual or Intended Pursuit of EHR Incentives . . . . .	149
8.5	Hospital Reasons for not Planning to Adopt EHRs . . . . .	151
8.6	Hospital Plans Regarding EHR Incentives . . . . .	152
9.1	Adoption of Household Products in the U.S. . . . .	158
9.2	Stock Prices for Two Leading EHRs . . . . .	167
E.1	Survival Curve . . . . .	191
E.2	Hazard Function . . . . .	192
F.1	Basic Confidence Interval Plots with Bootstrapping, 2008 Implementation . . . . .	195
F.2	Basic Confidence Interval Plots with Bootstrapping, 2010 Implementation . . . . .	196
F.3	Basic Confidence Interval Plots with Bootstrapping, 2011 Implementation . . . . .	197
G.1	REC RD Design Table - Multiple Models with Covariates . . .	199
G.2	REC RD design Table - Multiple Models with Covariates, Cut-off zoom . . . . .	200
I.1	MCD RD design Table with Covariates - Multiple Models . . .	204
I.2	MCD RD design Table with Covariates - Multiple Models, Zoom205	

# Chapter 1

## Introduction

Existing research has estimated the rates of electronic health record (EHR) adoption by physicians and hospitals in the U.S. and has identified some patterns in the rates of adoption. My research analyzes the preliminary impact of federal programs to promote the adoption of EHRs by estimating the impact of the programs on the patterns, rates, and levels of adoption.

In this study, I analyze the preliminary impact of several federal programs established to promote electronic health record (EHR) adoption on EHR adoption by estimating the impact of the programs on: (a) recognized disparities in adoption rates; (b) how long it will take to achieve widespread adoption; and (c) the level of adoption. I analyze survey data on the adoption of EHRs in conjunction with other data sets with provider or population level data. Through the Texas Health and Human Services Commission, EHR adoption surveys were fielded between 2010 and 2012 to identify the level of adoption and use of EHRs by health care practitioners and hospitals in Texas. Quantitative analysis based on these surveys is supplemented with qualitative analysis from key informant interviews with industry participants and stakeholders. Given the significant policy priority placed on EHR adoption, a better

understanding of the determinants and patterns of EHR adoption by health care providers can help to justify and guide the policies.

## 1.1 Questions and Hypotheses

*Top-level question:* What has been the preliminary impact of federal programs to promote the adoption of electronic health records?

*Sub-questions:*

1. Have the federal EHR adoption programs decreased disparities in the levels of adoption among different groups of providers?

Method: Apply logistic regression modeling to test whether providers with different characteristics are adopting EHRs at similar rates compared to estimates of the adoption rates of similar types of providers based on data collected at an earlier point in time.

Hypothesis: Some of the patterns of EHR adoption observed prior to the establishment of the federal EHR adoption programs are less pronounced in analyses of data collected after the establishment of the federal EHR adoption programs.

Theory: Several programmatic features of the federal EHR programs are specifically designed to address disparities in EHR adoption among providers. If these design features work the way they are intended, they will lessen some of the observed disparities.

Implication: If some of the patterns of EHR adoption are less pronounced



in analyses of data collected after the establishment of the federal EHR adoption programs, then it would suggest that one preliminary impact of the federal EHR adoption programs could be the reduction in certain disparities among providers associated with EHR adoption.

Alternate hypothesis/theory: Observed patterns of EHR adoption may remain largely unchanged. Despite programmatic features designed to address disparities in EHR adoption rates among providers, the federal EHR adoption programs may have a similar effect on adoption levels across-the-board, leaving patterns of EHR adoption largely unchanged.

2. How has the rate of EHR adoption changed since the establishment of the new federal EHR adoption programs?

Method: Apply a Bass technology diffusion model to project future levels of adoption and compare the results to similar analyses using data from prior to the establishment of the federal EHR adoption programs.

Hypothesis: Applying the Bass model to data that includes points in time after the establishment of the federal EHR adoption models shows a more rapid diffusion of EHRs into widespread use and a relative shift in the strength of external factors as opposed to internal factors in the diffusion.

Theory: The establishment of the federal EHR adoption programs is creating a more favorable environment for EHR adoption by lowering barriers to entry and increasing awareness.

Implication: If the projected rate of adoption is greater when estimated using data that includes points in time after the establishment of the federal EHR adoption programs, then it would suggest that one preliminary impact of the federal EHR adoption programs could be an increase in the rate of adoption.

Alternate hypothesis/theory: The rate of adoption is similar or lower than estimated prior to the establishment of the federal EHR adoption programs. Increased pressure from external sources (i.e., the federal EHR adoption programs themselves) induces resistance among providers, slowing a trend already underway in the absence of the programs.

3. Have the federal EHR promotion programs directly increased the level of EHR adoption among eligible professionals and eligible hospitals?

Method: Apply a regression discontinuity design to models of physician and hospital EHR adoption, to estimate the direct effects of the programs, utilizing programmatic design features of the EHR promotion programs as the cutoffs in the running variables. For the hospital model, use the 10% Medicaid patient volume threshold (an eligibility requirement for hospitals to receive the Medicaid EHR incentives) as the cutoff in the running variable. For the physician model, use the requirement that only physician groups of 10 or fewer can receive technical assistance from HIT Regional Extension Centers as the cutoff in the running variable.

Hypothesis: The Medicaid EHR adoption programs are having a statistically significant, but likely relatively small, impact on the level of EHR adoption among eligible hospitals.

Theory: The Medicaid EHR incentive payment program offers direct financial incentives for the adoption and use of EHRs. One of the recognized barriers to EHR adoption is cost, so subsidizing the purchase should increase the adoption level. The REC program offers direct technical support to physicians in small groups. Smaller groups are less likely to have the necessary technical wherewithal to approach the adoption of an EHR without assistance.

Implication: If the result of the regression discontinuity analysis suggests that the Medicaid EHR incentive payment program is estimated to be increasing the adoption level among eligible hospitals, then it would suggest that one preliminary impact of the federal EHR adoption could be increased levels of EHR adoption among eligible hospitals.

Alternate hypothesis/theory: The Medicaid EHR incentive payment program is not having a statistically significant impact on the adoption of EHRs by hospitals in Texas. Broader secular trends in the health care sector may be driving hospitals to adopt EHRs at substantially higher rates than historically, masking the possibly small marginal effect of the Medicaid EHR incentive payment program.

## **1.2 Policy Relevance**

Several federal programs have been established to try to increase the level of EHR adoption among physicians and hospitals. Significant federal funds are being disbursed through these programs. These programs include design elements that appear to be intended to address certain empirically known or theoretically anticipated barriers to EHR adoption and disparities in EHR adoption rates. This research estimates the impact these programs are having on long-term EHR adoption trends and estimates the immediate differential impact these programs are having on EHR adoption. In addition, this research estimates whether recognized disparities in adoption rates have been positively impacted by the federal EHR adoption programs and whether other disparities remain. This research informs policy development by suggesting how existing programs can be modified to increase adoption rates and address remaining disparities, and how programmatic elements of a potential new EHR adoption program could be designed to increase adoption rates and address remaining disparities.

## **1.3 Background**

In this background section, I discuss the clinical context for HIT, review HIT policy in general, and review EHR concepts in particular.

### **1.3.1 Description of HIT as a Supporting Technology for Clinical Practice and Health Services**

HIT comprises those information technologies used to support the delivery of health care at the point of care, and can broadly be divided into two categories – IT applications that allow for the automation of clinical support processes internal to an organization (e.g., electronic health records), and those IT applications that allow for the exchange of clinical information among different organizations (e.g., electronic prescribing, regional health information organizations, electronic lab results delivery; I refer to these applications collectively as health information exchange [HIE]).<sup>1</sup>

### **1.3.2 Electronic Health Records: What they are, how they are used, and why it matters**

EHRs are computer applications used to replace the medical record keeping function within an organization, whether that organization is a small health care practitioner’s office or a large hospital. The EHR market is fragmented with different EHR products having been developed for different uses, with the most significant variables being the clinical setting (e.g., inpatient vs. outpatient) and the specialty (e.g., primary care, specialty medicine).

The American Recovery and Reinvestment Act of 2009 (ARRA) included significant new policy and funding affecting EHRs including authorizing the Centers for Medicare and Medicaid Services (CMS) to establish programs

---

<sup>1</sup>*Defining Key Health Information Technology Terms* 2008.

through Medicare and state Medicaid programs to make incentive payments to eligible professionals (EPs) and eligible hospitals (EHs) for the meaningful use of certified EHRs. This legislation has given rise to significant federal rule-making activity and programmatic development through which a certification process has been structured and established, and "meaningful use" has been defined.

Implicit within the policy drive toward greater adoption of EHRs is a sense that their use will: (a) increase the administrative efficiency of provider organizations; (b) bring about greater adherence to clinical protocols and best practices through the use of clinical decision support systems; and (c) enable fundamental changes in the payment models used in the health care sector by making data on clinical processes and outcomes available to payers.

### **1.3.3 Appearance of HIT on the Policy Agenda**

Health information technology (HIT) as a public policy issue has moved through several distinct phases, from its initial appearance on the policy agenda to its emergence as a meaningful policy alternative, to its endorsement by Congress as a desired component of health care reform. Beginning in the late 1990s, it became apparent that the health care sector had been significantly slower than other industrial sectors at adopting and implementing information technology (IT) to support its core operations (i.e., the delivery of health care.) Around the same time, there were a number of high-profile reports describing systemic issues with the quality, safety, and efficiency of

health care in the U.S.<sup>2,3</sup> Also around the same time, macro-economic data began to show significant productivity gains in other industries that had been investing significantly in IT. These three perceived macro-trends in U.S. health care were widely considered to have been causally related, with the former contributing to the latter, leading to policies promoting greater adoption and use of health information technology (HIT).

#### **1.3.4 Emergence of HIT as a Meaningful Policy Alternative**

The recent history of high-profile federal programmatic activity and policy-making in support of HIT began with the establishment by President George W. Bush of the Office of the a National Coordinator for Health Information Technology (ONC) through Executive Order in 2004.<sup>4</sup> After a few years of rapid but fairly niche programmatic activity and policy development, the federal HIT program/policy agenda received a significant boost from the American Recovery and Reinvestment Act of 2009 (ARRA), which codified the ONC in Federal law, established several major new Federal HIT programs, and authorized significant new funds for HIT.<sup>5</sup> (As a minor historical aside, it may be useful to note that the HIT components of ARRA had previously been embodied as a "stand-alone" HIT bill known as the HIT for Economic and Clinical Health [HITECH] Act. This has led the HIT policy community to

---

<sup>2</sup>Kohn, Corrigan, and Donaldson 2000.

<sup>3</sup>Corrigan et al. 2001.

<sup>4</sup>Bush 2004.

<sup>5</sup>*The American Recovery and Reinvestment Act 2009.*

frequently refer to the HIT components of ARRA as the HITECH Act.)

### **1.3.5 Current State of Federal HIT Activity**

Beginning with federal executive-branch actions in 2004, boosted by the significant Congressional policy and funding through ARRA, beginning with its passage in 2009 and subsequent rulemaking, policy, and programmatic activity, the role of policy in promoting EHRs and HIE has been steadily increasing over the last decade. Currently, the federal government is promoting the adoption of EHRs and HIE through a coordinated suite of federal programs, including programs to establish a viable technical and legal framework for EHRs and HIE and programs to provide incentives for the adoption and use of EHRs and HIE.

In particular, the HIT components of ARRA included two major categories of HIT policy and funding: 1) EHR incentives; and 2) HIT infrastructure. The EHR incentive sections of ARRA authorized the establishment and funding of programs through Medicare and Medicaid to make incentive payments to eligible professionals (EPs) and eligible hospitals (EHs) for the meaningful use (MU) of certified electronic health record (EHR) systems. (Also included, with a separate incentive structure, were critical access hospitals (CAHs).) The HIT infrastructure sections of ARRA established several new programs including: the state health information exchange (HIE) program, for the establishment of state HIE systems/networks; HIT regional extension centers to support primary care physicians in the selection and implementation of



EHRs, and HIT workforce programs to educate workers on the implementation and administration of EHR and other HIT systems.

### **1.3.6 EHR Adoption Programs**

The Medicaid and Medicare EHR incentive payment programs and the HIT Regional Extension Center program, taken as a set, are all established to promote the adoption of EHRs among EPs and EHs. The category of EPs primarily refers to physicians for the Medicare incentive program, but also includes dentists, physician assistants under some circumstances, and nurse practitioners for the Medicaid incentive program. The Medicaid and Medicare EHR adoption programs include different eligibility and payment policies and apply these policies differently to EPs and EHs. As EPs, physicians must choose between participating in the Medicaid or Medicare EHR incentive payment programs. The Medicaid EHR incentive payment program requires for at least 30% of an EP's patient volume to be Medicaid and pays each EP a maximum of \$63,500 over 6 years. The Medicare EHR incentive payment program does not have a patient volume requirement, but does require EPs to be participating in the Medicare program, and pays each EP a maximum of \$44,000 over 6 years.

While EHs may participate in both the Medicare and Medicaid EHR incentive payment programs, their payment amounts are derived from a formula that takes into account the size of the hospital, but also the relative amounts of Medicaid and Medicare services provided by the hospital. The Medicare EHR

incentive payments for hospitals are paid evenly over four years. The Medicaid EHR incentive payments are paid over 3-5 years in proportions developed at state discretion. Hospital payments from the two programs can be several million dollars. In order to be eligible for the Medicaid EHR incentive payment program, at least 10% of a hospital's patient volume must be comprised of Medicaid patients. (Although not examined in this study, there is a different incentive payment calculation for critical access hospitals [CAHs], which are generally small, rural hospitals that meet certain additional criteria.)

In both the Medicaid and Medicare EHR incentive payment programs, for both EPs and EHRs, incentive payments are explicitly for the “meaningful use” of “certified electronic health records”. The statutory terms “meaningful use” and “certified electronic health record?” have been defined through the regulatory process. “Meaningful use” requires providers to demonstrate that they are using their EHRs in certain ways. The EHRs themselves must go through a federally structured certification process that is intended to ensure that they are capable of supporting providers to achieve “meaningful use”. The corresponding regulations structure the “meaningful use” and certification processes in three stages. The three stages of “meaningful use?” are intended to measure structures (i.e., that the appropriate mechanisms are in place to capture clinical activity as structured data), processes (i.e., clinical processes), and outcomes, respectively.

The certification process supports the “meaningful use” criteria by ensuring that the EHRs do not serve as an impediment to providers achieving the

“meaningful use” criteria. The certification process corresponding to stage one of “meaningful use?” largely tests the capability of EHRs to capture data in structured form. It is anticipated that the certification processes corresponding with later stages of “meaningful use” will test EHRs for a broader range of functionality including interoperability.

ARRA passed in early 2009, but there were significant regulatory activities that would need to be undertaken before the EHR promotion programs would be ready. CMS issued a Notice of Proposed Rulemaking (NPRM) for Stage 1 of Meaningful Use (MU) and ONC issued the corresponding NPRM for EHR certification in mid 2009, with a corresponding Interim Final Rule (IFR) published for each in late December of 2009. The final rules corresponding to each of these regulations were issued in late July 2010. The Stage 1 MU rule required for EPs and EHs to report, by attestation, to their achievement of the MU requirements. The MU requirements came in three primary primary categories – core objectives, menu set objectives, and clinical quality measures. When originally published as a final rule in mid 2010 (although somewhat modified in 2013), the MU requirements for EPs required that they meet 15 core objectives, 5 of the 10 menu set objectives, and report on 6 clinical quality measures. Each MU objective had a corresponding measure that had to be met in order for the EP or EH to get credit for achieving the objective. A representative objective/measure pair read as follows – “Objective: Implement drug-drug and drug-allergy interaction checks; Measure: The EP/EH/CAH has enabled this functionality for the entire EHR reporting pe-

riod”. In this case, the measure is a yes/no answer, but in many cases, the measure requires the EP or EH to achieve a certain level of performance. For example, another objective is to maintain an up-to-date problem list of current and active diagnoses, with the corresponding measure being that more than 80% of all unique patients seen by the EP or admitted to the EH or CAH have at least one entry or an indication that no problems are known for the patient recorded as structured data. Those objectives for which a percentage is the corresponding measure, required CMS to develop specific guidance regarding the calculation of the numerator and denominator that give rise to the percentage, including details regarding which cases can be excluded from the denominator (i.e., exclusion criteria.) For each MU measure, ONC and its advisory committees attempted to develop certification criteria for the EHRs that would ensure that the EHRs were capable of supporting the MU measure.

In order to receive the MU incentive payments, EPs and EHs had to attest to their achievement of the MU objectives, with measures calculated based on activities occurring during a 90 day period in the first year and for the whole year in each subsequent year. With the final rule for stage 1 MU and the corresponding certification rule, branded the “EHR Certification, 2011 Edition”, having been issued in mid 2010, the first period for which EPs and EHs could attest to MU was the first quarter of calendar year 2011. The first payments were received shortly thereafter, approximately mid-2011

At a high level, the Medicaid and Medicare programs operated fairly similarly, although for the first year of the Medicaid EHR incentive program,

providers, whether EPs or EOs only needed to attest that they had adopted, implemented, or upgraded a certified EHR system in order to receive payment. This criterion, known as AIU (adopt, implement, upgrade), applied to the first year of the EP payment for those EPs choosing to apply for the Medicaid incentives, and to the Medicaid component of the hospital payments, since most hospitals would likely be eligible for both Medicaid and Medicare incentive payments. Medicare incentive payments for EPs and the Medicare component of incentive payments for EOs required MU attestation in every year, including the first year.

In addition to the positive financial incentives for the “meaningful use” of certified EHRs, ARRA established penalties to be administered through the Medicare program for health care providers who do not use EHRs. Beginning in 2015, health care providers that do not use EHRs will have their Medicare reimbursements reduced by an increasing percentage. In 2015, Medicare reimbursements for health care providers who do not use EHRs will be cut by 1%, increasing to 2% in 2016 and 3% in 2017. The bill gives the Secretary of Health and Human Services the discretion to continue increasing the penalty amount by 1% per year up to a potential total penalty of 5% in 2019.

The HIT Regional Extension Center (REC) program was established through the health information infrastructure provisions of the HITECH Act to provide support to primary care physicians in small practices. ONC selected 65 nonprofit organizations throughout the country to serve as RECs. Each REC is responsible for assisting a certain number of “priority primary

care physicians” (PPCPs; basically PCPs practicing in small groups) to adopt EHRs and achieve “meaningful use”. Each REC has a target of at least 1,000 PPCPs and some have over 2,000. Nationally, all of the RECs are targeting a total of 100,000 PPCPs.

From the perspective of program evaluation, an important open question is that of the timing of the programmatic effect. Although the programs and funding were authorized in early 2009, the details of the programs were not promulgated through the agency program development processes until late 2009, and then particular grantees were not announced until early 2010. As I attempt to analyze the impacts of these programs on the adoption of EHRs in the physician and hospital sectors, I generally assume that the relevant program implementation date is early 2009, although I attempt to measure alternative timeframes through sensitivity analysis.

## **1.4 Contribution to the Literature**

The scope of research for this dissertation augments the existing literature and body of knowledge in both primary and secondary several distinct ways.

### **1.4.1 Primary contributions to the literature**

1. Analysis of patterns of EHR adoption Existing research has identified correlates of EHR adoption among physicians and hospitals. In many cases, these patterns of adoption correspond to anticipated barriers to

adoption. This research estimates changes in correlates of EHR adoption among physicians and hospitals using data collected in Texas, at two points in time, after the establishment of the federal EHR adoption programs and analyzed through logistic regression. This research also extends existing analyses of patterns of EHR adoption by testing additional potential explanatory variables. The patterns of EHR adoption thus derived are compared between two points in time as another measure of their impact.

2. Projections of future EHR adoption levels Existing research has projected how long it will take to achieve widespread EHR adoption by physicians by applying Technology Diffusion Models to data collected on levels of adoption at multiple points in time prior to the establishment of the federal EHR adoption programs. This research applies a Technology Diffusion Model to project future levels of EHR adoption using data collected in the Texas context after the establishment of the federal EHR adoption programs. The projections thus obtained are compared to projections based on data collected prior to the establishment of the federal EHR adoption programs as one measure of the impact of the programs.
3. Direct estimate of impact of federal EHR adoption program Through some convenient design features of two of the federal EHR promotion programs, it is possible to apply regression discontinuity (RD) designs to directly estimate the impact of the programs on the level of adoption

of EHRs by hospitals and the impact of the REC program on physician EHR adoption.

4. Qualitative validation of assumptions regarding EHR adoption dynamics

Through the use of semi-structured interviews, several aspects of EHR adoption are examined including anticipated barriers to adoption and the perceived impact of the federal EHR adoption programs.

### 1.4.2 Secondary contributions to the literature

1. Expands existing research on adoption rates. Existing research has measured the levels of EHR adoption by physicians and hospitals, in a number of different contexts, using varying methods. This study provides another perspective and data source within the existing body of research.
2. Supports post-implementation surveillance. The survey work analyzed herein provides an additional data point in the ongoing discussion of adoption rates.
3. Expands existing research on patterns of physician EHR adoption The existing literature on the relationship between physician practice characteristics and practice environment and relative rate of EHR adoption is limited. This research provides an additional data source and methodological approach to test existing hypotheses about these relationships, and expands on the range of practice environment factors tested.



## Chapter 2

### Literature Review

There has been a moderate amount of research on the rates of adoption and use of EHRs in the physician and hospital settings, including research on the impacts of EHR adoption and use. The existing literature can be categorized into quadrants based on this characterization (physician or hospital setting on one axis, adoption rates or impacts on the other). Notable across all of these quadrants in the existing literature is a lack of convergence. There is a significant variance in reported adoption rates of EHRs, even when the same populations are surveyed. There is also little agreement on the impact of EHRs on clinical outcomes – so little agreement, in fact, that even the direction of the effect remains a point of significant debate.

#### 2.1 Physician EHR Adoption

Numerous studies have been performed and published on the rate of EHR adoption by physicians in many different practice locations, and across many different specialties. Surveys have been done on the adoption of EHRs in

community health centers,<sup>1</sup> in the general ambulatory care environment,<sup>2,3,4,5,6,7,8,9</sup> as well as in multiple specialties including geriatrics,<sup>10</sup> obstetrics/gynecology,<sup>11,12</sup> pediatrics,<sup>13</sup> and primary care.<sup>14,15,16,17</sup> These surveys are also generally done at either the national level or a sub-state level, leaving significant room for additional survey work and analysis of state-level data for the purpose of policy development and programmatic guidance. Of particular note for this research, there have been several state-level surveys of EHR adoption by physicians in Texas by the Texas Medical Association, but they suffer from methodological flaws of uncertain impact.<sup>18,19,20,21</sup> While a number of surveys of EHR adoption by physicians have been performed, there is a notable lack of convergence among the findings with some surveys finding rates of adoption 2 to 3 times the rates of other surveys. (See table 2.1 on page 21.)

---

<sup>1</sup>Shields et al. 2007.

<sup>2</sup>Bates 2005.

<sup>3</sup>Simon et al. 2007.

<sup>4</sup>Simon et al. 2008.

<sup>5</sup>DesRoches et al. 2008.

<sup>6</sup>Hsiao 2008.

<sup>7</sup>Hsiao and Hing 2012.

<sup>8</sup>Furukawa et al. 2014.

<sup>9</sup>DesRoches 2015.

<sup>10</sup>Yeager, Menachemi, and Brooks 2010.

<sup>11</sup>Lagrew, Stutman, and Sicaeros 2008.

<sup>12</sup>Raglan et al. 2014.

<sup>13</sup>Lehmann et al. 2015.

<sup>14</sup>Andrews et al. 2004.

<sup>15</sup>Bazemore et al. 2011.

<sup>16</sup>Stream 2009.

<sup>17</sup>Xierali et al. 2013.

<sup>18</sup>*TMA Special Survey Electronic Medical Records Report* 2008.

<sup>19</sup>*TMA Special Survey Electronic Medical Records Report* 2009.

<sup>20</sup>*TMA Special Survey Electronic Medical Records Report* 2010.

<sup>21</sup>*TMA Special Survey Electronic Medical Records Report* 2012.

Sample of "Top-line" Results from Select EHR Adoption Surveys				
	National			Texas
	DesRoches et al	NCHS (Hsiao et al)	Shields et al	Texas Medical Association
Year	2008	2008	2007	2007
Adoption Rate	13%	20%	13%	33%

Table 2.1: EHR Adoption Level

## 2.2 Hospital EHR Adoption

There has also been a moderate amount of research on EHR adoption by hospitals in the U.S., including broad-based surveys,<sup>22,23,24,25</sup> state-level surveys, research focused on particular types of hospitals (most frequently children's hospitals),<sup>26,27,28</sup> and research focused on particular hospital functions or departments (most frequently the emergency department).<sup>29,30,31</sup> Hospital EHRs are generally significantly more complex than EHRs in the ambulatory care environment because they cover so many different clinical functions across the many different departments and provider types. Electronic systems in fully "paperless" hospital environments must be able to document the

---

<sup>22</sup>Jha et al. 2009b.

<sup>23</sup>Jha et al. 2010.

<sup>24</sup>DesRoches et al. 2013.

<sup>25</sup>Adler-Milstein et al. 2014.

<sup>26</sup>Menachemi et al. 2009.

<sup>27</sup>Nakamura et al. 2010.

<sup>28</sup>Teufel et al. 2013.

<sup>29</sup>Geisler, Schuur, and Pallin 2010.

<sup>30</sup>Pallin et al. 2010.

<sup>31</sup>Jamoom and Hing 2015.

clinical information captured in each department and by each provider type and support the many lines of communication among providers. Thus, as a practical matter, hospitals rarely implement electronic systems across all departments at one time. As an empirical matter, most hospitals have some electronic clinical systems, but relatively few are fully paperless. Therefore, surveys of EHR adoption by hospitals often seek to establish adoption of EHRs by hospitals along a continuum rather than as a binary value. Of the broad-based surveys performed to date, those fielded by Jha et al and published in mid 2009 and late 2010 are methodologically strong and probably represent the most accurate baseline of hospital EHR adoption. According to Jha et al (2009 and 2010), the percentage of U.S. hospitals that had adopted basic or comprehensive EHRs in 2008 was approximately 8.7%, rising to 11.9% in 2009.

### **2.3 Determinants of EHR Adoption**

There is a modest literature on determinants of EHR adoption by physicians. The primary categories of factors generally identified as being predictive of EHR adoption include practice characteristics,<sup>32,33,34,35,36,37</sup> market charac-

---

<sup>32</sup>Burt and Sisk 2005.

<sup>33</sup>Simon et al. 2007.

<sup>34</sup>Bramble et al. 2010.

<sup>35</sup>Bramble et al. 2010.

<sup>36</sup>Menachemi, Powers, and Brooks 2011.

<sup>37</sup>Grinspan et al. 2013.

teristics,<sup>38,39,40,41</sup> and patient population.<sup>42,43</sup> There is little consensus on the determinants of EHR adoption, but common factors evaluated (and in some but not all analyses found to be significant) include practice size, Medicare population, employment by an HMO, and age of physician. (Although Texas currently has few physicians employed by HMOs, this characteristic has been tested in analysis of national data.) There is also a somewhat more limited literature on the determinants of hospital EHR adoption, which includes size<sup>44,45</sup>, efficiency,<sup>46</sup> and economic characteristics<sup>47,48</sup>.

## 2.4 Impacts of EHR Adoption

Existing evaluations of the impact of EHR adoption by hospitals has generally focused on the impact on quality of care, across a range of metrics.<sup>49,50,51,52</sup> Generally, these studies have found no impact or relatively small relationships between EHR adoption and performance on quality metrics. Al-

---

<sup>38</sup>Abdolrasulnia et al. 2008.

<sup>39</sup>Menachemi et al. 2007.

<sup>40</sup>Menachemi et al. 2012.

<sup>41</sup>Menachemi et al. 2012.

<sup>42</sup>Jha et al. 2009a.

<sup>43</sup>Hing and Burt 2009.

<sup>44</sup>Jha et al. 2009b.

<sup>45</sup>Zhang et al. 2013.

<sup>46</sup>Zhivan and Diana 2012.

<sup>47</sup>Shin et al. 2012.

<sup>48</sup>Ginn, Shen, and Moseley 2011.

<sup>49</sup>DesRoches et al. 2010.

<sup>50</sup>Elnahal et al. 2011.

<sup>51</sup>Jones et al. 2010.

<sup>52</sup>McCullough et al. 2010.

though efficiency as an outcome measure might also seem to be a likely direction for evaluation, efficiency has been more often evaluated as a potential determinant of hospital EHR adoption.<sup>53</sup> There is, in addition to the literature on the impacts of hospital EHR adoption, also a somewhat more modest literature on the impacts of physician EHR adoption – more limited likely because so much less physician-level outcome data is collected than hospital-level outcome data. There have been several meta-analyses of other research on the impacts of HIT on different outcome measures.<sup>54,55,56</sup> These meta-analyses have identified an increasing trend in positive findings from analyses of lots of miscellaneous HIT-related interventions in numerous different contexts. There has been very little published research on the impacts of EHR use by physicians.

## 2.5 Diffusion of Innovations

The increasing adoption of a new technology within a community or market is often described as the diffusion of innovation. Originally popularized almost half a century ago, diffusion of innovation remains an active area of academic study and applied theory. In *Diffusion of Innovations*, first published in 1962, Everett Rogers, a rural sociologist, developed a broad theory that describes how technological innovations are adopted and spread across

---

<sup>53</sup>Zhivan and Diana 2012.

<sup>54</sup>Chaudhry et al. 2006.

<sup>55</sup>Goldzweig et al. 2009.

<sup>56</sup>Buntin et al. 2011.

populations.<sup>57</sup> Rogers reviewed over 500 studies of technology adoption and developed a theoretical framework describing many aspects of the diffusion of innovation. Key aspects of Rogers' theory include the elements of diffusion and the categories of adopters. According to Rogers, the four primary factors that affect how a new innovation will spread throughout a population are innovation, communications channels, time, and a social system. The categories of adopters as defined by Rogers are innovators, early adopters, early majority, late majority, and laggards. Diffusion of Innovations remains influential and timely, with the fifth edition having been published in 2003.<sup>58</sup> The theory laid out in *Diffusion of Innovations* was rendered into a quantitative form by Frank Bass in the late 1960s.<sup>59</sup> The Bass diffusion model was originally used, and continues frequently to be used to model the adoption of new consumer products. The original Bass article is the fourth most cited paper ever published in the journal *Management Science*. The EHR adoption literature includes the use of the Bass diffusion model to project future levels of EHR adoption.<sup>60,61</sup>

---

<sup>57</sup>Rogers 1962.

<sup>58</sup>Rogers 2003.

<sup>59</sup>Bass 1969.

<sup>60</sup>Ford, Menachemi, and Phillips 2006.

<sup>61</sup>Ford et al. 2009.

## Chapter 3

### Data

The primary sources of data for this research are surveys of EHR adoption by physicians and hospitals conducted in Texas and key informant interviews.

#### 3.1 Surveys

Through the Office of e-Health Coordination, which I managed at the Texas Health and Human Services Commission, I supervised the development and administration of two surveys of EHR adoption and use. My staff administered the hospital survey, which we sent to all of the licensed hospitals in Texas. The hospital survey was sent initially by email, with email and phone follow-up. The practitioner survey was developed under my supervision and administered under contract by the Public Policy Research Institute (PPRI) at Texas A&M University. The practitioner survey sample was a random group of practitioners identified from licensing data by PPRI.

The hospital surveys were fielded in 2010 and 2012 by staff from the Office of e-Health Coordination at the Texas Health and Human Services Commission. In both years, all hospitals in the state were contacted for potential



inclusion in the survey. In 2010, there were 253 completed hospital surveys from the 583 licensed hospitals in that year, for a 43% response rate, and in 2012, there were 177 completed hospital surveys from the 580 licensed, operating hospitals in that year, for a 31% response rate.

The first practitioner survey was fielded in 2011. The sampling frame included 7,430 physicians in Texas, and resulted in 1,239 completed responses for a response rate of 16.7%. Surveys were fielded primarily by mail, with follow-up by phone, and subsequent follow-up, at the request of the physician, by email or fax. The survey was fielded between February 18 and April 28, 2011.

The second practitioner survey was fielded in 2012. The sampling frame included 7,428 physicians – the same physicians in the 2011 survey, minus two who were no longer considered practicing anymore, either due to lack of licensure, relocation, or death. The 2012 survey resulted in 923 completed responses for a response rate of 12.4%. As in 2011, surveys were fielded primarily by mail, with follow-up by phone, and subsequent follow-up, at the request of the physician, by email or fax. The survey was fielded between June 21 and October 12, 2012.

The practitioner surveys did not include identifying information at the practitioner level. The hospital data did identify the hospitals by name, which allows for some additional analysis of changes by hospital and additional supplementation based on new questions in the 2012 survey.

In addition to asking direct questions about EHR use, the practitioner survey also asks a number of questions about whether specific clinical and administrative functions within the practitioner's practice are done electronically. The data from these questions can be used to provide some context for the range of EHR adoption rates reported across various surveys.

<b>Primary Data Sources</b>				
<i>Survey</i>	<i>Unit of Analysis</i>	<i>N</i>	<i>Description</i>	<i>Date</i>
Texas Hospital EHR Adoption Survey (2010)	Hospital	253	Survey of Texas hospitals performed by Texas HHSC, Office of e-Health Coordination	2010
Texas Hospital EHR Adoption Survey (2012)	Hospital	177	Survey of Texas hospitals performed by Texas HHSC, Office of e-Health Coordination	2012
Texas Practitioner EHR Adoption Survey (2011)	Individual practitioner	1,239	Survey of Texas practitioners (physicians, dentists, physician assistants, nurse practitioners) performed by Texas A&M Public Policy Research Institute under contract with Texas HHSC Office of e-Health Coordination	2011
Texas Practitioner EHR Adoption Survey (2012)	Individual practitioner	923	Survey of Texas practitioners (physicians, dentists, physician assistants, nurse practitioners) performed by Texas A&M Public Policy Research Institute under contract with Texas HHSC Office of e-Health Coordination	2012

Table 3.1: Primary Data Sources

<b>Supplemental Data Sources</b>				
<i>Data Source</i>	<i>Unit of Analysis</i>	<i>N</i>	<i>Description</i>	<i>Date</i>
Estimates of the Total Populations of Counties and Places in Texas	Counties and Places	NA	Derived by the Texas State Data Center from multiple data sources.	2010
American Community Survey	Household	2.9 million	Annually collected by the U.S. Census Bureau. Includes data on race and ethnicity, and poverty level by zip code.	2012
Sizes of Texas Counties	County	254	Geographic sizes of Texas counties.	NA
Dartmouth Atlas	Hospital Referral Regions	22	Geographic regions corresponding with local health care markets.	2012
Urban/Rural Classification by County	County	254	U.S. Department of Health and Human Services, Health Resources and Services Administration.	2012

Table 3.2: Supplemental Data Sources

## 3.2 Data Representativeness

Due to the relatively low response rates, particularly for the physician survey, it is important to test for non-response bias to determine if the sample appears to be representative of the population. As was noted in the literature review and as is considered in more detail in the quantitative analysis, one of the characteristics most commonly identified as correlating with EHR adoption for both physicians and hospitals is size. Therefore, I compare the distribution of size of physician practices and hospitals in the survey samples against population estimates to determine if the samples appear to be representative. Put another way, the comparisons of the size distributions in the samples to the estimated size distributions in the population helps me to understand the potential role of non-response bias in the data.

Figure 3.1 on page 32 shows a comparison of the distribution of the size groups in which physicians practice from the practitioner surveys used in this analysis to the distribution of of the size groups in which physicians practice in the 2012 edition of the biannual survey conducted by the Texas Medical Association. As shown in the figure, the proportion of physicians practicing in small and large practices in the sample are slightly smaller than the proportion of physicians practicing in small and large practices in the population as estimated by the TMA survey. In addition, the proportion of physicians practicing in medium sized practices in the survey is substantially greater than in the population, and the proportion of physicians practicing in very large practices in the survey is substantially less than in the population.

All population estimates are based on the TMA biannual physician survey. It is also important to note that the unit of analysis in both the practitioner survey used in this analysis and the TMA survey is the practitioner; that is, the estimates do not reflect the number of physician groups of the different sizes, but rather the number of physicians practicing in groups of different sizes. The differences in proportions across the physician practice size categories is considered as part of the interpretation of the quantitative analysis.

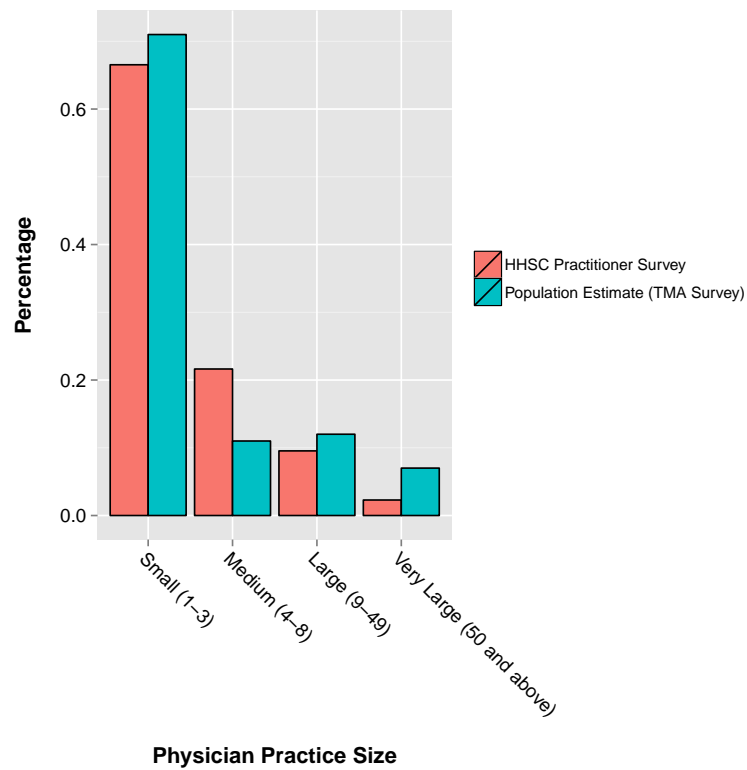


Figure 3.1: Comparison of Physician Practice Size Distribution

Figure 3.2 on page 34 shows a comparison of the distributions of proportions across several hospital size groups for the hospital survey used in this analysis and for the population as estimated by the annual hospital survey conducted by the Texas Department of State Health Services (DSHS). DSHS serves as the state regulator of hospitals and requires participation in the survey; as such, the estimates obtained from the DSHS survey can reasonably be considered true population measures. The proportions in the size categories from the hospital survey vary from the proportions in the size categories in the population. The proportion of hospitals in the small category is smaller in the sample than in the population, while the proportion of hospitals in the large category is larger in the sample than in the population. As in the case of the physician survey, the differences in proportions across the hospital categories is considered as part of the interpretation of the quantitative analysis.

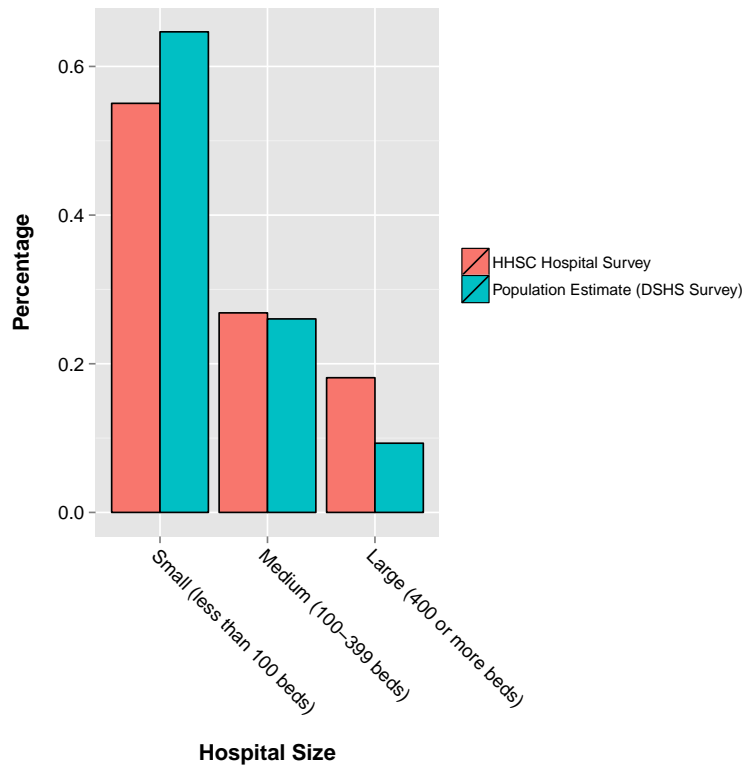


Figure 3.2: Comparison of Hospital Size Distribution



### **3.3 Key Informant Interviews**

Key informant interviews were performed with two classes of physician outliers. Physicians who were predicted by an early version of the logit model to have the greatest likelihood of adopting an EHR, but who had not were interviewed. Physicians who were predicted by an early version of the logit model to have the least likelihood of adopting, but who had adopted, were also interviewed. The purpose of interviewing these groups was to better understand the adoption dynamics by physicians whose predicted and actual behavior diverged the most.

# Chapter 4

## Methodology

This section describes how I developed and analyzed the data, potential methodological issues identified ex ante, and the approach to using the key informant interviews to complement the analysis of survey data.

### 4.1 Data Set Development

In the subsequent analysis, I link the data from the provider surveys to other data sources to identify potential determinants and patterns of adoption. The practitioner survey includes location data that I use to link it with socio-economic and demographic data through which I am able to test whether the socio-economic character or rurality of a practitioner's practice environment impacts the rate of EHR adoption (i.e., is there a digital divide?) The practitioner survey itself also includes data on practice size and payer mix that is used to test other potential determinants of EHR adoption. The hospital survey includes the names of the hospitals and is linked to other data on hospital characteristics.

Data Set	Hospital correlates
Purpose	Identify correlates of hospital EHR adoption.
Component Data Sources	<ul style="list-style-type: none"> <li>• Texas Hospital EHR Adoption Survey</li> <li>• County characteristics</li> <li>• Zipcode characteristics</li> <li>• Hospital referral region characteristics</li> <li>• Urban/rural categorization</li> </ul>
Link	Hospital name/ID
Questions Addressed	What characteristics of hospitals' environments (especially socio-economic and geographic) affect their rates of EHR adoption?

Table 4.1: Hospital Data Set

Data Set	Physician correlates
Purpose	Identify regional/ practice-setting characteristics correlated with physician EHR adoption
Component Data Sources	<ul style="list-style-type: none"> <li>• Texas Practitioner EHR Adoption Survey</li> <li>• Estimates of the Total Populations of Counties and Places in Texas</li> <li>• County characteristics</li> <li>• Zipcode characteristics</li> <li>• Hospital referral region characteristics</li> <li>• Urban/rural categorization</li> </ul>
Link	County/ Zip code
Questions Addressed	<ul style="list-style-type: none"> <li>• What characteristics of physicians practices (e.g., number of providers in practice, affiliations with larger organizations) affect their rates of EHR adoption?</li> <li>• What characteristics of physicians practice environments (especially socio-economic and geographic) affect their rates of EHR adoption?</li> </ul>

Table 4.2: Physician Data Set

## 4.2 Quantitative Models

In this study, patterns of EHR adoption are estimated using logistic regression models, long-term rates and levels of EHR adoption are projected using a Technology Diffusion Model (TDM), and the direct impact of two of the federal EHR adoption programs are estimated using regression discontinuity designs.

### 4.2.1 Logistic Regression Models

In this study, I analyze the determinants of physician and hospital EHR adoption using logistic regression models. The modeling begins with the standard logistic regression model. The literature suggests a number of potential correlates of EHR adoption including patient socio-economics, geography (primarily urban vs. rural), and payer-mix. In the subsequent analysis, I model each of these possibilities separately and together. Results thus obtained are compared to EHR adoption patterns observed in the literature.

### 4.2.2 Bass Diffusion Model

The Bass diffusion model can be represented by an equation of the following form:

$$F(t) = (1 - e^{-(p+q)t}) / (1 + (\frac{q}{p})e^{-(p+q)t})$$

In this study, I develop a data set of physician EHR adoption rates from multiple surveys and fit the curve to the data through the application of a linear optimization routine, as has been done in the published literature on

EHR adoption. The data set includes the Texas physician survey discussed above. The estimated  $p$  and  $q$  values are then compared to  $p$  and  $q$  values from estimates using data only prior to the establishment of the federal EHR programs. Tracking the Bass exposition, changes in the relative values of  $p$  and  $q$  can be interpreted as a shift in the relative strength of internal versus external pressures to adopt the technology. The estimated  $p$  and  $q$  values are also used to project levels of adoption and specific dates in the future and to compare these levels with projections based on  $p$  and  $q$  values derived from data collected only prior to the establishment of the federal EHR programs.

The structure of the Bass diffusion function would seem to make the shape of the curve fairly sensitive to relatively small variations in the underlying data, but little discussion of this point has been included in the published literature on EHR adoption that utilizes this approach. The methodological analysis includes sensitivity analysis to test this possibility

### **4.2.3 Regression Discontinuity Design**

Programmatic features of both the Health IT Regional Extension Center Program, which provides direct technical assistance to physicians in small groups, and the financial incentives for the meaningful use of EHRs for hospitals through the Medicaid program, allow for analysis through a regression discontinuity (RD) design. For each of the analyses leveraging the RD design, a programmatic feature establishes a sharp cut-off in eligibility for services or incentives.

For the Health IT Regional Extension Center program, only primary care physicians practicing in groups of 10 or fewer are eligible. Physician practice size is a data element in the physician surveys.

For the Medicaid EHR incentive payment program, a hospital must have a patient volume of at least 10% Medicaid in order to be eligible for the program. The hospital survey includes hospitals with all levels of Medicaid patient volume and questions about Medicaid patient volume. Since the literature suggests that payer-mix is correlated with EHR adoption rates, there may be some endogeneity issues associated with using a value of Medicaid patient volume for the cutoff. However, since Medicaid patient volume is a continuous variable and nothing about the particular cutoff value would seem to systematically affect the EHR adoption decision, it may not make a difference, but the possibility will be tested.

#### **4.2.4 Statistical Software**

The R statistics platform was used for all statistical analyses conducted herein. All of the R code used to generate the analysis, tables, and figures contained in this study can be found in Appendix K.

## Chapter 5

# Have patterns in EHR adoption been changing in ways consistent with the programmatic features of the federal EHR promotion programs?

### 5.1 Overview

Several major federal programs to promote the adoption of EHRs by practitioners and hospitals were authorized by Congress in early 2009 and have been ongoing since then. Chief among these are direct financial incentives for the "meaningful use" of certified EHRs and the establishment of a network of HIT Regional Extension Centers to help physicians in small practices to adopt EHRs and become meaningful users of EHRs. These programs were established based, in part, on findings from the literature identifying financial barriers as an impediment to EHR adoption and use; and practice size as a common correlate of EHR adoption and use. The particular features of these programs appear to be designed to address these issues directly.

Using the data collected from 2010-2012 on practitioner and hospital EHR adoption and use, I demonstrate whether these policies and programs have helped to alleviate the particular impediments to EHR adoption previously identified. In addition, I will try to determine whether there might be



other factors that might be driving disparities in EHR adoption and use that might reasonably be addressed through different programs/policies or changes to programmatic features in existing programs.

In the following analysis, I analyze the correlation between a number of different factors and EHR adoption and use by physicians and hospitals and then discuss them in the context of the existing EHR policies and programs and compare those findings with related findings from the literature. The primary analytic approach used in this portion of the analysis is logistic (logit) regression, with EHR adoption as dependent variable and miscellaneous other characteristics as the independent variables.

Although a number of different types of practitioners might use EHRs, and all of the practitioner types surveyed by the practitioner survey (physicians, dentists, NPs, and PAs) can be eligible for EHR incentives under the federally funded programs, only physicians are targeted by the HIT Regional Extension Centers. Therefore, for the purposes of the analysis in this section, the practitioner data analyzed are limited to responses from physicians. The literature on EHR adoption by physicians suggests that rurality, practice size, payer mix, patient demographics, and physician age may all be correlated with EHR adoption. The literature on hospital EHR adoption suggests that rurality, size, status as a teaching hospital, and ownership status may all be correlated with EHR adoption.

## 5.2 Key Questions

The key question motivating this chapter is whether the patterns of EHR adoption, by physicians and hospitals, are changing in ways that are consistent with the programmatic features of the federal programs to promote the adoption of EHRs. Specifically, the federal EHR promotion programs targeted at physicians included the Regional Extension Center program, which provided direct technical assistance to primary care physicians practicing in small groups, and a substantially higher set of financial incentives for physicians providing a significant amount of care to Medicaid enrollees. On the hospital side, the federal programs provided some specific additional support to Regional Extension Centers for supporting Critical Access Hospitals, which are a particular class of rural hospital; and a sliding-scale of financial incentives that increased with greater provision of services to Medicare and Medicaid enrollees.

## 5.3 Hypotheses

Although the two years of data come from consecutive years, they represent two points in time soon after the establishment of the federal programs. As such, it is reasonable to suggest that the difference between these points in time represents the pace of change in key characteristics of EHR adoption, across both physicians and hospitals. The key method that will be deployed in this chapter is logistic regression, through which it will be possible to estimate the relative contributions to EHR adoption from a number of other charac-

teristics of physicians and hospitals. With the inclusion of data from both the 2011 and 2012 surveys, it is possible to identify variations in the relative contributions of the other characteristics by including year interaction terms in the logit regressions. These coefficient estimates associated with these year interaction terms will be the key points of analysis to determine whether the patterns of adoption changed between 2011 and 2012. Identification of changes in the patterns of adoption between these two years, even if such changes coincide with features of the federal programs to promote the adoption of EHRs, will only be able to be interpreted as consistent with the programmatic goals, not caused by the programmatic features.

## **5.4 Model Development**

Generally, the characteristics most likely to affect EHR adoption by physicians and hospitals are those that directly or indirectly describe the financial resources, technical capacity, and local market dynamics. Access to greater financial resources will likely be correlated with higher rates of EHR adoption since cost could be a factor inhibiting adoption. Technical capacity includes the presence of technical ability and the flexibility to apply that ability, but also willingness to use technology more generally. Local market dynamics could affect EHR adoption either by creating peer pressure to adopt, pressure from certain dominant payers or other institutions, or competitive dynamics that lead to substantially lower or higher margins than in otherwise similar markets.

Consistent with this perspective, characteristics identified in the literature as relating to EHR adoption include whether the provider is a Medicaid provider, the size of the provider (for physicians, practice size; for hospitals, bed count), metropolitan status, and socio-economic character of the area. Direct or proxy measures for these characteristics are tested in the following analyses. In addition to logit regression, the subsequent models are tested for robustness with linear probability models.

## **5.5 Physician EHR Adoption Patterns**

### **5.5.1 Data**

The practitioner survey was fielded in 2011 and 2012 by Texas A&M Public Policy Research Institute under contract with the Office of e-Health Coordination within the Texas Health and Human Services Commission. The purposes of the survey was to identify the levels of adoption and use of health information technology among Texas physicians, with particular focus on the levels of adoption of electronic health records. The practitioner surveys were fielded by fax and phone using a sample of practitioners randomly selected from the corresponding lists of licensed practitioners. Given the interest in trying to identify the effect of the federal programs authorized in 2009, for which full implementation did not occur until 2010, the physician portion of this analysis will focus on the more recent year of data.

### **5.5.1.1 Survey**

The first practitioner survey was fielded in 2011. The sampling frame included 7,430 physicians in Texas, and resulted in 1,239 completed responses for a response rate of 16.7%. Surveys were fielded primarily by mail, with follow-up by phone, and subsequent follow-up, at the request of the physician, by email or fax. The survey was fielded between February 18 and April 28, 2011.

The second practitioner survey was fielded in 2012. The sampling frame included 7,428 physicians – the same physicians in the 2011 survey, minus two who were no longer considered practicing anymore, either due to lack of licensure, relocation, or death. The 2012 survey resulted in 923 completed responses for a response rate of 12.4%. As in 2011, surveys were fielded primarily by mail, with follow-up by phone, and subsequent follow-up, at the request of the physician, by email or fax. The survey was fielded between June 21 and October 12, 2012.

### **5.5.1.2 Data Set Construction**

The primary data set used for the physician analysis is developed from the combination of the 2011 and 2012 practitioner surveys, combined with demographic and hospital referral pattern data at the zipcode level. The 2011 and 2012 data are merged and, in subsequent logistic regression analysis, year fixed effects are estimated.

### 5.5.1.3 Key Variables

For the practitioner survey, the following categories of EHR adoption were captured:

1. EHR is implemented for all practitioners and all practice locations (paperless)
2. EHR is implemented for some practitioners or in some locations (partially paperless)
3. Implementation is planned in the next year
4. Implementation is planned in the next two years
5. Implementation is not planned

Figure 5.1 on page 49 shows the distribution of answers to the practitioner survey for the top-line EHR adoption question before re-coding for 2012 practitioner survey. For the purposes of the logit models, both full and partial EHR adoption were coded as 1 and all other answers were coded as 0. Forty-six percent of responding physicians indicated that they had adopted a full EHR and another 13 percent indicated partial adoption. After re-coding in preparation for further analysis, nearly 60 percent of physician responses were coded for EHR adoption.

Table 5.1 on page 50 shows summary statistics for a number of the key variables in the subsequent logit analyses. About 67 percent of respondents

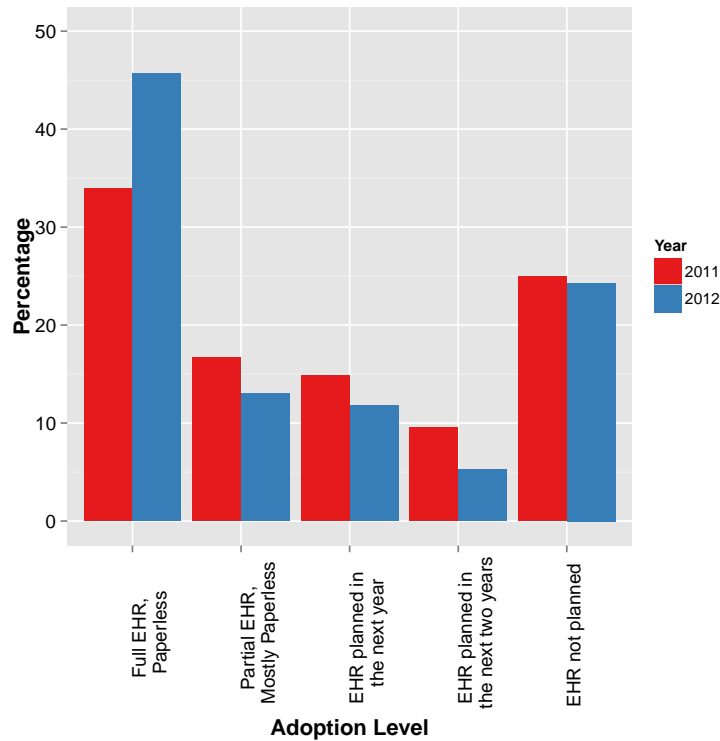


Figure 5.1: EHR Adoption Level by Year

indicated that they participated in the Medicaid program and about 54 percent reported EHR adoption. The distribution of physician practice sizes is very skewed toward the origin. Almost 80% of the physician respondents reported practicing in a solo or small group (2-5 physicians) setting.

About 50 percent of the responding physicians were in large central metro areas, about 16 percent were in large fringe metro areas, about 12 percent were in medium metro areas, about 13 percent were in small metro areas, about 6 percent were in micropolitan areas, and about 4 percent were in non-

<b>Variable</b>	<b>Levels</b>	<b>n<sub>0</sub></b>	<b>%<sub>0</sub></b>	<b>n<sub>1</sub></b>	<b>%<sub>1</sub></b>	<b>n<sub>all</sub></b>	<b>%<sub>all</sub></b>
EHR Adoption	0	536	49.4	325	41.4	861	46.0
	1	550	50.6	461	58.6	1011	54.0
	all	1086	100.0	786	100.0	1872	100.0
Medicaid Participation	0	328	30.2	284	36.1	612	32.7
	1	758	69.8	502	63.9	1260	67.3
	all	1086	100.0	786	100.0	1872	100.0
Practice Size	Solo (1)	429	39.5	323	41.1	752	40.2
	Small (2-5)	429	39.5	305	38.9	734	39.2
	Medium (6-10)	127	11.7	92	11.7	219	11.7
	Large (11-100)	89	8.2	57	7.3	146	7.8
	Very Large (101-1000)	12	1.1	8	1.0	20	1.1
	all	1086	100.0	785	100.0	1871	100.0
Percent Poverty by Zipcode	Less than 10	269	24.9	215	27.4	484	26.0
	Between 10 and 20	405	37.5	286	36.5	691	37.1
	Between 20 and 30	230	21.3	153	19.5	383	20.6
	Between 30 and 40	126	11.7	93	11.9	219	11.8
	Between 40 and 50	29	2.7	21	2.7	50	2.7
	Between 50 and 60	3	0.3	4	0.5	7	0.4
	Between 60 and 70	17	1.6	12	1.5	29	1.6
	all	1079	100.0	784	100.0	1863	100.0
County Metro Status	Noncore	40	3.7	32	4.1	72	3.8
	Micropolitan	77	7.1	40	5.1	117	6.2
	Small metro	143	13.2	108	13.7	251	13.4
	Medium metro	131	12.1	88	11.2	219	11.7
	Large fringe metro	167	15.4	124	15.8	291	15.5
	Large central metro	528	48.6	394	50.1	922	49.2
	all	1086	100.0	786	100.0	1872	100.0

Table 5.1: Summary Table

core areas. Poverty percentages were identified at the zipcode level, and ranged from 0 percent to about 66 percent. For the analysis using categorical variables, the poverty percentage measures were grouped into bins 10 percentage points wide. More than half of responding physicians were in areas with less than 20 percent poverty, and only about 5 percent were in areas with more than 40 percent poverty.



## 5.5.2 Results

### 5.5.2.1 Bivariate Analysis

A number of the variables examined herein exhibit strong bivariate correlations with the EHR adoption variable and with other independent variables.

Figure 5.2 on page 52 is a correlogram that provides a quick overview of pairwise correlations among the continuous variables in the physician analysis. Pairwise correlations between the continuous variables and discrete variables are shown in subsequent tables. Population density, physician group size, and poverty percent by zipcode show slight positive correlations, with little obvious pattern to the shapes of the correlation distributions.



Figure 5.2: Physician Variable Correlogram

The following tables show other relevant bivariate correlations in the models.

Table 5.2 on page 54 shows the cross-tabulation between EHR adoption and the categorical versions of the other dependent variables. There are several noteworthy findings from the cross-tabulation. While it might be assumed that physicians seeing Medicaid patients have access to a lower level of financial, and thus technical resources, a higher percentage of EHR adopters are Medicaid providers than otherwise. As has been well-documented in the literature, there appears to be a slight trend among EHR adopters toward larger practices. Less expected, but consistent with the observation regarding Medicaid participation, is a slight trend among EHR adopters toward areas with higher rates of poverty. Finally, there appears also to be a slight trend among EHR adopters toward smaller, more rural areas.

<b>Variable</b>	<b>Levels</b>	<b>n<sub>0</sub></b>	<b>%<sub>0</sub></b>	<b>n<sub>1</sub></b>	<b>%<sub>1</sub></b>	<b>n<sub>all</sub></b>	<b>%<sub>all</sub></b>
Medicaid Participation	0	324	37.6	288	28.5	612	32.7
	1	537	62.4	723	71.5	1260	67.3
	all	861	100.0	1011	100.0	1872	100.0
Practice Size	Solo (1)	447	52.0	305	30.2	752	40.2
	Small (2-5)	310	36.0	424	41.9	734	39.2
	Medium (6-10)	69	8.0	150	14.8	219	11.7
	Large (11-100)	32	3.7	114	11.3	146	7.8
	Very Large (101-1000)	2	0.2	18	1.8	20	1.1
	all	860	100.0	1011	100.0	1871	100.0
Percent Poverty by Zipcode	Less than 10	242	28.1	242	24.1	484	26.0
	Between 10 and 20	302	35.1	389	38.8	691	37.1
	Between 20 and 30	181	21.1	202	20.1	383	20.6
	Between 30 and 40	91	10.6	128	12.8	219	11.8
	Between 40 and 50	22	2.6	28	2.8	50	2.7
	Between 50 and 60	4	0.5	3	0.3	7	0.4
	Between 60 and 70	18	2.1	11	1.1	29	1.6
	all	860	100.0	1003	100.0	1863	100.0
County Metro Status	Noncore	25	2.9	47	4.6	72	3.8
	Micropolitan	72	8.4	45	4.4	117	6.2
	Small metro	92	10.7	159	15.7	251	13.4
	Medium metro	89	10.3	130	12.9	219	11.7
	Large fringe metro	129	15.0	162	16.0	291	15.5
	Large central metro	454	52.7	468	46.3	922	49.2
	all	861	100.0	1011	100.0	1872	100.0

Table 5.2: Cross-tabs by EHR Adoption

Table 5.3 on page 55 shows the cross-tabulation between Medicaid participation and the categorical versions of the other dependent variables.

<b>Variable</b>	<b>Levels</b>	<b>n<sub>0</sub></b>	<b>%<sub>0</sub></b>	<b>n<sub>1</sub></b>	<b>%<sub>1</sub></b>	<b>n<sub>all</sub></b>	<b>%<sub>all</sub></b>
EHR Adoption	0	324	52.9	537	42.6	861	46.0
	1	288	47.1	723	57.4	1011	54.0
	all	612	100.0	1260	100.0	1872	100.0
Practice Size	Solo (1)	276	45.2	476	37.8	752	40.2
	Small (2-5)	223	36.5	511	40.6	734	39.2
	Medium (6-10)	60	9.8	159	12.6	219	11.7
	Large (11-100)	44	7.2	102	8.1	146	7.8
	Very Large (101-1000)	8	1.3	12	0.9	20	1.1
	all	611	100.0	1260	100.0	1871	100.0
Percent Poverty by Zipcode	Less than 10	249	40.8	235	18.8	484	26.0
	Between 10 and 20	201	33.0	490	39.1	691	37.1
	Between 20 and 30	74	12.1	309	24.7	383	20.6
	Between 30 and 40	61	10.0	158	12.6	219	11.8
	Between 40 and 50	13	2.1	37	3.0	50	2.7
	Between 50 and 60	0	0.0	7	0.6	7	0.4
	Between 60 and 70	12	2.0	17	1.4	29	1.6
	all	610	100.0	1253	100.0	1863	100.0
County Metro Status	Noncore	5	0.8	67	5.3	72	3.8
	Micropolitan	16	2.6	101	8.0	117	6.2
	Small metro	31	5.1	220	17.5	251	13.4
	Medium metro	29	4.7	190	15.1	219	11.7
	Large fringe metro	123	20.1	168	13.3	291	15.5
	Large central metro	408	66.7	514	40.8	922	49.2
	all	612	100.0	1260	100.0	1872	100.0

Table 5.3: Cross-tabs by Medicaid Participation

In addition to Medicaid providers being higher adopters of EHRs than non-Medicaid providers, they are also tend to practice in small and medium practices rather than solo practices, tend to practice in areas with higher rates of poverty, and are more frequently practicing outside of large urban areas than non-Medicaid providers.

### 5.5.2.2 Logistic Analysis

Based on the data sets thus constructed, several logit models of EHR adoption by physicians are developed. These models are clustered into two sets. The first set (table 5.4 on page 57) maintains a number of the independent variables as continuous variables, whereas in the second set, with results displayed across two tables (table B.1 on page 180 for the non-interacted explanatory variables and table B.2 on page 181 for the interaction terms), the independent variables are all re-coded into categorical variables. For all of these models, data is included from both the 2011 and 2012 physician surveys. Year fixed effects and interaction terms are also included.

In the first set (table 5.4 on page 57), the first model for physician EHR adoption includes whether the physician is a Medicaid provider, and the practice size. The second model adds poverty percentage, as measured at the zipcode level, as a measure of relative population wealth (or lack thereof). Finally, the third physician model adds a variable representing the population density of the county in order to assess whether some population densities might have a significant impact on the model.

In these models, status as a Medicaid provider and practice size are statistically significant, across all three models. The fixed effects coefficient representing the second year of the survey (2012) is moderately significant in the first model, but not in subsequent models. Finally, the interaction term between practice size and year is significant across all three models, but negative and of a similar size to the positive value for the model as a

whole, suggesting that the positive, statistically significant effect shown in the coefficients for the non-interacted practice size terms across the three models, should only be interpreted to apply to 2011.

Several "Pseudo-R-Squared" values and other measures of model strength are included in table 5.5 on page 58. Measures of logit model strength are notoriously quirky, so should be used with some caution, but they can be useful for measuring the relative strength of different models using the same dependent variable and data. In this case, the goodness-of-fit measures do not provide much guidance on judging among the models.

	<i>Dependent variable:</i>		
	EHR		
	(1)	(2)	(3)
Medicaid Provider	0.37***	0.38***	0.36***
Practice Size	0.01**	0.01**	0.01**
Poverty Pct (by zipcode)		-0.002	-0.002
Population Density			-0.0000
Survey Year	0.29*	0.30	0.35
Medicaid Provider*Year	0.18	0.18	0.16
Practice Size*Year	-0.01*	-0.01*	-0.01*
Poverty Pct (by zipcode)*Year		-0.001	-0.0003
Population Density*Year			-0.0000
Constant	-0.31***	-0.28**	-0.22
Observations	1,872	1,864	1,864
Log Likelihood	-1,265.43	-1,262.52	-1,261.78
Akaike Inf. Crit.	2,542.87	2,541.04	2,543.55

*Note:*

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Table 5.4: Physicians, Both Years, Continuous Models

	McFadden	Cox.Snell	Nagelkerke	Count	AIC
Model 1	0.0202	0.0275	0.0368	0.5737	2542.8688
Model 2	0.0187	0.0255	0.0341	0.5687	2541.0365
Model 3	0.0193	0.0263	0.0351	0.5746	2543.5506

Table 5.5: Additional Summary Statistics, Continuous Models



	<i>Dependent variable:</i>		
	EHR		
	(1)	(2)	(3)
Medicaid Provider	0.32**	0.29**	0.24
Practice Size - Small	0.68***	0.67***	0.65***
Practice Size - Medium	1.05***	1.04***	1.07***
Practice Size - Large	1.62***	1.61***	1.64***
Practice Size - Very Large	2.99***	2.75***	2.73**
Poverty Pct (10-20)		0.09	0.15
Poverty Pct (20-30)		-0.04	0.05
Poverty Pct (30-40)		0.16	0.14
Poverty Pct (40-50)		0.38	0.47
Poverty Pct (50-60)		-1.41	-1.19
Poverty Pct (60-70)		-0.77	-0.66
Metro-Micro			-1.43***
Metro-Small			-0.52
Metro-Medium			-0.33
Metro-Large Fringe			-0.54
Metro-Large Central			
Observations	1,871	1,862	1,862
Log Likelihood	-1,211.58	-1,204.10	-1,188.66
Akaike Inf. Crit.	2,447.15	2,456.20	2,445.32

*Note:*

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Table 5.6: Physicians, Both Years, Factor Models, Explanatory Variables

	<i>Dependent variable:</i>		
	EHR		
	(1)	(2)	(3)
Survey Year			-0.64
Medicaid Provider*Year			0.26
Practice Size - Small*Year			0.03
Practice Size - Medium*Year			0.23
Practice Size - Large*Year	0.14	0.25	0.27
Practice Size - Very Large*Year	-0.91	-1.07	-1.14
Poverty Pct (10-20)*Year		0.01	0.06
Poverty Pct (20-30)*Year		-0.08	-0.002
Poverty Pct (30-40)*Year		-0.03	0.22
Poverty Pct (40-50)*Year		-0.96	-0.91
Poverty Pct (50-60)*Year		0.43	0.35
Poverty Pct (60-70)*Year		0.33	0.43
Metro-Micro*Year			1.04
Metro-Small*Year			1.11*
Metro-Medium*Year			0.24
Metro-Large Fringe*Year			1.10*
Metro-Large Central*Year			
Observations	1,871	1,862	1,862
Log Likelihood	-1,211.58	-1,204.10	-1,188.66
Akaike Inf. Crit.	2,447.15	2,456.20	2,445.32

*Note:* \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Table 5.7: Physicians, Both Years, Factor Models (cont.), Year Effects

	McFadden	Cox.Snell	Nagelkerke	Count	AIC
Model 1	0.0614	0.0812	0.1085	0.6077	2447.1523
Model 2	0.0630	0.0833	0.1113	0.6122	2456.2000
Model 3	0.0750	0.0984	0.1314	0.6224	2445.3192

Table 5.8: Additional Summary Statistics, Factor Models

In the second set of models, with results shown in table B.1 on page 180 for the non-interacted explanatory variables and table B.2 on page 181 for the interaction terms, the continuous variables for practice size and poverty percent are restructured as categorical variables. The practice size variable is broken into five categories – Solo (1 physician), Small (2-5), Medium (6-10), Large (11-100), and Extra Large (101-1000). The poverty percent variable is broken into evenly-spaced blocks, 10 percentage points wide, across the range of values (0%-66%). The coefficients for these variables are estimated independently and are estimated relative to the values representing solo physician and the lowest poverty range. The variable representing rurality is coded into 6 levels following one of the common federal urban/rural scales. The coefficient estimates for metro status are all independently relative to the large urban category, which is the omitted dummy variable. The Medicaid provider variable is coded as in the previous set of models.

In this second set of models, Medicaid status remains highly significant, and all of the coefficients representing different practice sizes are also highly significant, with similar effect sizes across the models. Several of the rural categories, which are measured at the county level and are relative to

counties with a non-core value, also show statistical significance. In particular, physicians practicing in micro counties and large central counties showed statistically significantly lower level of EHR adoption, while the year interaction terms for physicians practicing in small and large fringe showed a positive effect at a low level of statistical significance.

As in the case of the first set of models, several "Pseudo-R-Squared" values and other measures of model strength are included in table 5.8 on page 61. Unlike the models using continuous variables, these models appear to continue to improve as additional variables are added, with progressively improving scores on most of the measures of model strength.

In addition to the models described above, an additional set of models was tested that included dummy variables for the different medical trading areas in Texas, using the Hospital Referral Regions (HRRs) as developed for the Dartmouth Atlas for Healthcare. The results from these models were very similar to the results of the models without using the HRRs, with significant coefficient estimates, at a low level of significance for only one of the 22 HRRs. Results for these models are included in Appendix A for reference. The robustness of the models is also tested using linear probability models, which are included in Appendix B. All significance levels associated with coefficient estimates in the linear probability models are the same as those calculated for coefficient estimates in the logit models.

### 5.5.3 Hypothesis Test

The motivating question for this chapter is whether the patterns in EHR adoption are changing since the establishment of the federal programs to promote the adoption of EHRs. The relevance of the question is due to the fact that certain features of those programs are specifically targeted at improving adoption rates among small physician practices and those serving Medicaid recipients. As such, the variables of greatest interest are those reflecting status as a Medicaid provider and practice size.

The effect of status as a Medicaid provider is statistically significant across most of the models discussed above, although not in the direction expected. Certain features of the federal EHR programs to promote the adoption of EHRs seem to be designed to provide larger incentives to Medicaid providers than other providers, thus suggesting that Medicaid providers are perceived to adopt EHRs at lower rates than other providers. While this is consistent with the “digital divide” rhetoric, the data analyzed herein do not bear out that conclusion. To the contrary, in the models described above, being a Medicaid provider is positively correlated with EHR adoption at a statistically significant level in almost every model tested

The other characteristic consistently identified as statistically significant in the models considered above, is practice size. In both the continuous and categorical sets of models, practice size is consistently positively correlated with EHR adoption and statistically significant. Also notable is that, in the categorical models, the effect sizes (all relative to solo practice physicians),

increase across every size grouping of physicians, across each model tested.

Considering these results formally, particularly in light of the question of whether the patterns of adoption are changing, requires particular focus on the year interaction terms. For the continuous models, the coefficient estimate for the year fixed effect variable was slightly statistically significant in the simplest model tested, but not significant for the other models. The other statistically significant coefficient estimate among the year interaction terms for the continuous models was the practice size, which was negative and consistently statistically significant at a relatively low level of significance. The most obvious interpretation of this finding is that the effect of practice size on adoption rates was less in 2012 than in 2011.

Turning our attention to the models with categorical dependent variables, the coefficient estimate for the year fixed effect variable is not significant in any of the models and the statistically significant year effects for practice size no longer appear. There are a couple of notable results among the coefficient estimates for the rurality categories (small and large fringe counties), both positive and significant at a low level. While this could suggest an increase in adoption levels in these category types, the federal EHR promotion programs did not have any programmatic characteristics specifically targeted at rural counties as such, making the interpretation of these effects of limited relevance for the present inquiry.

#### **5.5.4 Discussion**

The most robust results from these analyses were that status as a Medicaid provider and practice size are positively correlated with EHR adoption. Past analysis and commentary on the issue of physician EHR adoption has occasionally highlighted the potential of a "digital divide" wherein physicians in areas with high rates of poverty and/or low incomes might be less likely to adopt EHRs. The analyses discussed herein did not identify any significant differences in the adoption rates of EHRs by physicians in areas with different levels of poverty. The key question framing this chapter has been – "Have federal programs to promote the adoption of electronic health records mitigated disparities in adoption patterns?" One of the most significant disparities in physician EHR adoption patterns noted in the previous literature was by practice size. This pattern appears still to have been very much in place in 2011, with some suggestion that the effect may have been somewhat less dramatic in 2012.

#### **5.5.5 Limitations**

This analysis has structural and statistical limitations. While this study is not able to support strong causal claims regarding the effects of the EHR promotion programs on the patterns of EHR adoption, it can identify changes in the patterns of EHR adoption between the two years. In addition, the response rate was low (overall, below 20%), which could lead to significant self-selection or other bias. Additionally, survey questions regarding the adop-

tion of EHRs are notoriously fragile – the particular forms of the question have a significant impact on the response, making comparison of results across different survey instruments and data collection methods problematic. Finally, the two surveys were only taken in 2011 and 2012. In addition to being temporally close together, thus potentially not providing enough time for changes in EHR adoption patterns to be recognized, these years are also relatively close to the beginning of the EHR promotion programs; if the effects of the programs are lagged or increase over time, then they may not be seen yet, or only partially, in 2011 and 2012.

#### **5.5.6 Conclusion**

Some of the features of the federal EHR promotion programs are structured to address specific perceived disparities in EHR adoption between certain groups of physicians. In particular, the Regional Extension Center program provides direct technical assistance to physicians practicing in groups of 10 or fewer and the financial incentives for the meaningful use of EHRs provide are substantially larger for physicians seeing a large percentage of Medicaid patients. The question examined in this section was whether the patterns of EHR adoption changed between 2011 and 2012 in ways that were consistent with those structural goals of the EHR promotion programs. The results identified herein do not support dismissing the relevant null hypothesis (i.e, that the patterns of physician EHR adoption are not changing in ways consistent with the goals implicit in the federal EHR promotion programs).



## **5.6 Hospital EHR Adoption Patterns**

### **5.6.1 Data**

The primary data sets used for this portion of the analysis are based on the 2011 and 2012 survey data collected on hospital adoption and use of health information technology in Texas, as described in the data section. For this analysis, the data from the two surveys are pooled and a year variable is added.

#### **5.6.1.1 Survey**

Through the Office of e-Health Coordination, which I managed at the Texas Health and Human Services Commission, I supervised the development and administration of two surveys of EHR adoption and use. My staff administered the hospital survey, which we sent to all of the licensed hospitals in Texas. The hospital survey was sent initially by email, with email and phone follow-up. The practitioner survey was developed under my supervision and administered under contract by the Public Policy Research Institute (PPRI) at Texas A&M University. The practitioner survey sample was a random group of practitioners identified from licensing data by PPRI.

The hospital surveys were fielded in 2010 and 2012 by staff from the Office of e-Health Coordination at the Texas Health and Human Services Commission. In both years, all hospitals in the state were contacted for potential inclusion in the survey. In 2010, there were 253 completed hospital surveys from the 583 licensed hospitals in that year, for a 43% response rate, and in

2012, there were 177 completed hospital surveys from the 580 licensed, operating hospitals in that year, for a 31% response rate.

Given the interest in trying to identify the effect of the federal programs authorized in 2009, for which full implementation did not occur until 2010, the hospital portion of this analysis will focus on the more recent year of data.

### 5.6.1.2 Data Set Construction

The primary data sets used for the hospital analysis are developed from the combination of the 2010 and 2012 hospital surveys with hospital-level case-load and volume data from the Texas Department of State Health Services' Annual Hospital Survey, payer-data from the Texas Health Care Information Collection's hospital discharge data, county-level demographic data, and zip-code-level data on hospital referral patterns and poverty.

### 5.6.1.3 Key Variables

Statistic	Mean	St. Dev.	Min	Max
Total Beds	185.868	279.399	0	1,674
Total Revenue (millions)	134.721	275.604	0	2,790
Population Density	715.012	907.271	0.629	2,605.586
Poverty % by Zip	19.364	10.595	2.500	66.700
Charity Ratio	0.194	1.792	-0.00000	34.602

Table 5.9: Hospital Data, Summary Statistics, Continuous Variables

Table 5.9 on page 68 provides summary statistics for a number of the continuous variables used in the following analysis. While qualitatively sim-

<b>Variable</b>	<b>Levels</b>	<b>n<sub>0</sub></b>	<b>%<sub>0</sub></b>	<b>n<sub>1</sub></b>	<b>%<sub>1</sub></b>	<b>n<sub>all</sub></b>	<b>%<sub>all</sub></b>
EHR Adoption	0	137	58.8	45	30.2	182	47.6
	1	96	41.2	104	69.8	200	52.4
	all	233	100.0	149	100.0	382	100.0
Hospital Size	Small (0-100)	145	63.0	82	55.0	227	59.9
	Medium (101-400)	59	25.6	40	26.9	99	26.1
	Large (400+)	26	11.3	27	18.1	53	14.0
	all	230	100.0	149	100.0	379	100.0
County Metro Status	Large central metro	73	31.3	44	29.5	117	30.6
	Large fringe metro	30	12.9	23	15.4	53	13.9
	Medium metro	23	9.9	14	9.4	37	9.7
	Small metro	35	15.0	24	16.1	59	15.4
	Micropolitan	24	10.3	17	11.4	41	10.7
	Noncore	48	20.6	27	18.1	75	19.6
	all	233	100.0	149	100.0	382	100.0
Percent Poverty by Zipcode	Less than 10	40	17.2	25	16.9	65	17.1
	Between 10 and 20	90	38.8	55	37.2	145	38.2
	Between 20 and 30	76	32.8	45	30.4	121	31.8
	Between 30 and 40	18	7.8	18	12.2	36	9.5
	Between 40 and 50	4	1.7	4	2.7	8	2.1
	More than 50	4	1.7	1	0.7	5	1.3
all	232	100.0	148	100.0	380	100.0	
Survey Year	0	233	100.0	0	0.0	233	61.0
	1	0	0.0	149	100.0	149	39.0
	all	233	100.0	149	100.0	382	100.0

Table 5.10: Hospital Data, Summary Statistics, Nominal Variables

ilar to the physician practice size distribution in its skew toward the origin, the hospital size distribution does not drop off as abruptly, moving up in size. A number of other variables relating to volume of services were analyzed for potential inclusion, including inpatient volume, ED visits, and Medicare days. These other variables were examined to determine whether they perhaps might reflect other, independent characteristics, but all were found to be highly correlated with total beds and total revenue and were not included in subsequent analysis, nor are the included in this table. It appears that they are all rough proxies for size/volume.

Table 5.10 on page 69 provides summary statistics for some of the key categorical variables included in the hospital analysis. Most of the hospitals in the sample are in the small category and in metropolitan areas. The distribution in poverty percentage across the responding hospitals appears roughly to follow a normal distribution.

## **5.6.2 Results**

### **5.6.2.1 Bivariate Analysis**

A number of the independent variables involved in the preceding logit models exhibit strong bivariate correlations with the EHR adoption variable and with other independent variables.

The following correlogram provide a quick overview of correlations between the continuous variables in the hospital data set by showing qualitative shadings in the lower left and scatterplots in the upper right. As in the bivariate analysis of physician data, correlations in the hospital data between discrete variables will be shown with cross-tabulations, and correlations between discrete variables and continuous variables will be shown with subsequent boxplots.

Figure 5.3 on page 72 is a correlogram of the variables included in this analysis. The similar distributions among many of the variables noted in the discussion of key variables above is also reflected in the correlogram with all of the variables that seem most likely to scale with size showing strong bivariate correlations. As in the physician data, median income and poverty percent

show a strong negative correlation. The socio-economic variables show very little correlation with the size, charity care, or population density variables.

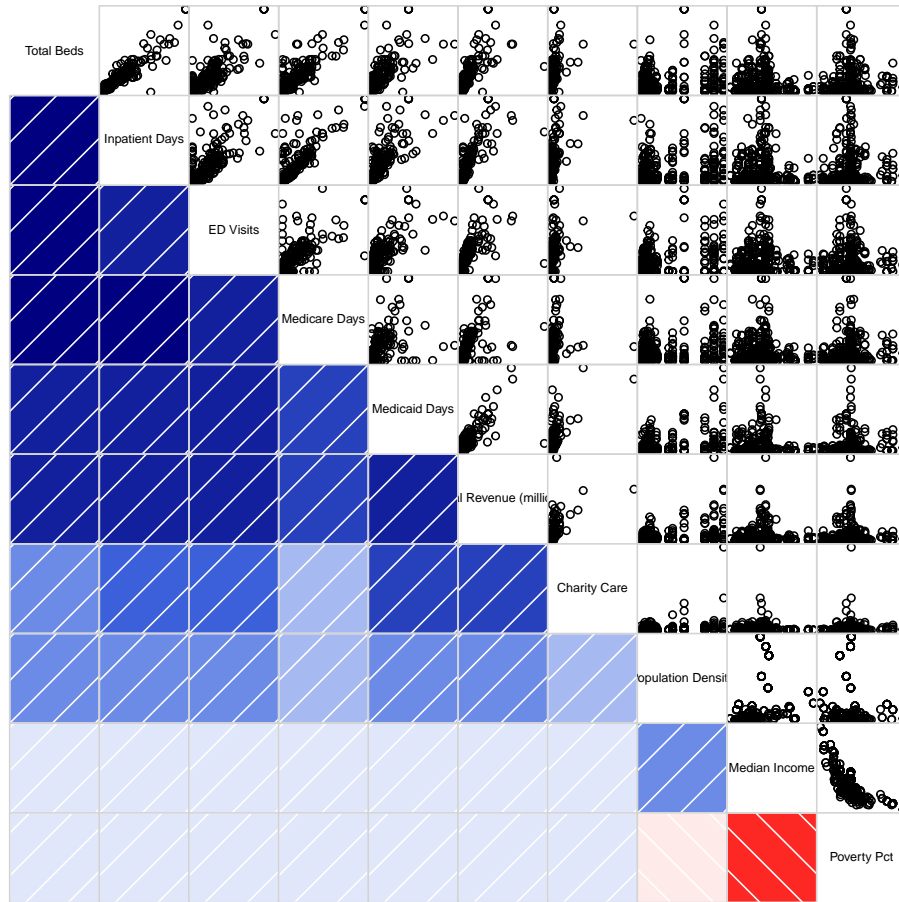


Figure 5.3: Hospital Variables Correlogram, 2012

<b>Variable</b>	<b>Levels</b>	<b>n<sub>0</sub></b>	<b>%<sub>0</sub></b>	<b>n<sub>1</sub></b>	<b>%<sub>1</sub></b>	<b>n<sub>all</sub></b>	<b>%<sub>all</sub></b>
Hospital Size	Small (0-100)	131	72.8	96	48.2	227	59.9
	Medium (101-400)	39	21.7	60	30.1	99	26.1
	Large (400+)	10	5.6	43	21.6	53	14.0
	all	180	100.0	199	100.0	379	100.0
County Metro Status	Large central metro	48	26.4	69	34.5	117	30.6
	Large fringe metro	29	15.9	24	12.0	53	13.9
	Medium metro	16	8.8	21	10.5	37	9.7
	Small metro	35	19.2	24	12.0	59	15.4
	Micropolitan	20	11.0	21	10.5	41	10.7
	Noncore	34	18.7	41	20.5	75	19.6
	all	182	100.0	200	100.0	382	100.0
Percent Poverty by Zipcode	Less than 10	40	22.0	25	12.6	65	17.1
	Between 10 and 20	63	34.6	82	41.4	145	38.2
	Between 20 and 30	62	34.1	59	29.8	121	31.8
	Between 30 and 40	10	5.5	26	13.1	36	9.5
	Between 40 and 50	2	1.1	6	3.0	8	2.1
	More than 50	5	2.8	0	0.0	5	1.3
	all	182	100.0	198	100.0	380	100.0

Table 5.11: Hospital Data, Cross-tabs by EHR Adoption

Table 5.11 on page 73 shows the cross-tabulation values for EHR adoption and a number of the other categorical variables. Hospitals that have adopted EHRs are disproportionately in the large and medium categories and in more urban counties. Although the data does not show a consistent pattern, hospitals that have adopted EHRs appear to be slightly more likely to be in areas with higher poverty than non-adopting hospitals.

### 5.6.2.2 Logistic Analysis

The hospital models follow a similar pattern and logic to the physician models, with some differences based on data availability and operational differences. Data on patient volume by payer and other financial information come from the Department of State Health Services' hospital discharge data,

population density is calculated from county-level population and geographic data, and income-related variables are from county-level demographic data.

Table 5.12 on page 75 shows the first set of hospital EHR adoption logit models, which are composed only of continuous variables. This set of models uses continuous variables. The first model in this set includes only size-related variables, the second adds a geographic variable, and the third adds socio-economic variables. In these models, the only coefficient estimate that is significant across all three models is total revenue, and it is only significant at a low level. The coefficient estimate for the year fixed effect variable is highly significant and positive across the first two models, but not significant in the last model, while none of the coefficient estimates for any of the individual year interaction variables are significant. These results would seem to indicate that the patterns of adoption did not change significantly between 2010 and 2012, but the rate of adoption did change significantly.

Additional summary statistics are shown in table 5.13 on page 76. The several goodness-of-fit estimates generally show an increasing model strength across the models.

Table D.1 on page 188 and table D.2 on page 189 show the second set of hospital EHR adoption logit models, using categorical variables. The first part of each model, with the non-interacted explanatory variables is in the first table, and the second part of each model with the year effect interaction terms is in the second table. Similar to the first set of hospital models, the first model in this set includes only size-related variables, the second adds a



	<i>Dependent variable:</i>		
	ehr		
	(1)	(2)	(3)
Total Beds	0.001	0.001	0.0005
Total Revenue (millions)	0.004*	0.004*	0.004*
Population Density		-0.0001	-0.0001
Poverty Pct			-0.020
Charity Ratio			1.894
Survey Year	1.149***	1.024***	0.614
Total Beds*Year	-0.001	-0.0005	-0.002
Total Revenue (millions)*Year	0.001	-0.00001	0.001
Population Density*Year		0.0003	0.0002
Poverty Pct*Year			0.023
Charity Ratio*Year			0.142
Constant	-0.767***	-0.718***	-0.480
Observations	380	380	376
Log Likelihood	-229.853	-229.420	-224.040
Akaike Inf. Crit.	471.705	474.839	472.081

*Note:*

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Table 5.12: Hospitals, Both Years, Continuous Models

	McFadden	Cox.Snell	Nagelkerke	Count	AIC
Model 1	0.1256	0.1595	0.2129	0.6553	471.7052
Model 2	0.1272	0.1614	0.2154	0.6579	474.8393
Model 3	0.1392	0.1753	0.2339	0.6596	472.0809

Table 5.13: "Goodness of Fit" Statistics, Continuous Models

<i>Dependent variable:</i>			
ehr			
	(1)	(2)	(3)
Size - Medium (100-400)	0.87***	1.00***	1.02***
Size - Large (more than 400)	1.58***	1.81***	2.10***
Metro-Large Fringe		0.10	-0.04
Metro-Medium		0.27	0.11
Metro-Small		-0.53	-0.66
Metro-Micro		0.64	0.54
Metro-Noncore		0.45	0.34
Poverty pct (10-20)			0.31
Poverty (20-30)			-0.08
Poverty pct (30-40)			0.03
Poverty pct (40-50)			-0.37
Poverty pct (more than 50)			
Observations	379	379	377
Log Likelihood	-232.96	-222.48	-211.35
Akaike Inf. Crit.	477.92	476.95	474.70

*Note:* \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Table 5.14: Hospitals, Both Years, Factor Models, Explanatory Variables

geographic variable, and the third adds a socio-economic variable.

Given the strong bivariate correlations between the size-related variables, all except the total bed count have been removed in this set, and total

	<i>Dependent variable:</i>		
	ehr		
	(1)	(2)	(3)
Survey Year			0.86
Size - Medium (100-400)*Year			0.15
Size - Large (more than 400)*Year			0.13
Metro-Large Fringe*Year			-0.84
Metro-Medium*Year		-0.74	-0.59
Metro-Small*Year		-0.45	0.07
Metro-Micro*Year		-2.00**	-1.96**
Metro-Noncore*Year		0.63	0.82
Poverty pct (10-20)*Year			1.08
Poverty pct (20-30)*Year			0.17
Poverty pct (30-40)*Year			1.27
Poverty pct (40-50)*Year			16.15
Poverty pct (more than 50)*Year			0.36
Constant			
Observations	379	379	377
Log Likelihood	-232.96	-222.48	-211.35
Akaike Inf. Crit.	477.92	476.95	474.70

*Note:* \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Table 5.15: Hospitals, Both Years, Factor Models (cont), Year Effects

	McFadden	Cox.Snell	Nagelkerke	Count	AIC
Model 1	0.1116	0.1431	0.1910	0.6649	477.9242
Model 2	0.1516	0.1892	0.2525	0.6649	476.9508
Model 3	0.1900	0.2313	0.3086	0.7003	474.7009

Table 5.16: "Goodness of Fit" Statistics - Factor Models

bed count has been re-coded into small, medium, and large categories, with cutoffs 100 and 400. The baseline for the size category is the small hospital category (less than 100 beds). The metropolitan status categorical measure is based on HRSA urban-rural codes linked to the hospital counties. The baseline for the metropolitan status measure is 'Metro-Large Core'. The poverty percent values by zipcode span the range of 2.5%-66.7%. The poverty percent categories are recoded from the continuous poverty percent variable and are assigned to bins ten percentage points wide, ranging up to 50%. The baseline for the poverty percent category is 'less than 10%'.

In this set of models, all of the coefficient estimates related to the non-interacted size variables are positive and highly significant, with direction and scale within expectations. No other coefficient estimates associated with the non-interacted variables are statistically significant. As in the case of the first set of hospital models, the coefficient estimate for the year effect is statistically significant in the first two models, but not in the third. Of the remaining interacted variables, only the coefficient estimates corresponding to the micro-sized counties is significant. It is difficult to know how to interpret this estimate since the coefficient estimates for county metropolitan status, in both the non-interacted and interacted variables, are not monotonically changing along with

the county metropolitan status.

Additional summary statistics are shown in 5.16 on page 78. As in the first set of models, the goodness-of-fit values show an increasing model strength across the models.

In addition to the models described above, an additional set of models was tested that included dummy variables for the different medical trading areas in Texas, using the Hospital Referral Regions as developed for the Dartmouth Atlas for Healthcare. These models did not result in any new significant coefficient estimates. Results for these models are included in Appendix C for reference. The robustness of the models is also tested using linear probability models, which are included in Appendix D. The only different significance calculation between the models is for the coefficient estimate representing areas with poverty percentage greater than 50%. The coefficient estimate in the linear probability model is moderately significant ( $p < 0.05$ ), whereas in the logit model, it was not significant.

### 5.6.3 Hypothesis Test

The key question motivating this chapter has been – “Have patterns in EHR adoption been changing in ways consistent with the programmatic features of the federal EHR promotion programs?” One of the features of the hospital component of the EHR incentive programs is the scaling of the amount of the incentive based on amount of Medicaid and Medicare volume. As such, particularly with respect to the added financial incentive due to

higher Medicaid volume, it could be reasonable to anticipate that hospitals serving higher proportions of Medicaid patients, which would likely correlate with those in areas of higher poverty, would be more likely to adopt EHRs. In the results, this could have appeared as a positive significant coefficient estimate corresponding to one or more of the variables representing poverty percentage (continuous or categorical) interacted with the year. This result did not appear in the models, so we cannot dismiss the null hypothesis that the patterns of hospital EHR adoption relating to location in areas with higher poverty did not change.

#### **5.6.4 Discussion**

In the bivariate analysis, the most consistent and notable result was that hospitals adopting EHRs seemed to be characterized by higher average values on a number of variables that would generally trend with the size of the institution. Although this result did not appear in a significant way in the initial set of logit models using continuous variables, it did re-emerge in the second set of models, using categorical variables. Variables associated with geographic and socio-economic characteristics were not consistently significant across the models. Previous research has shown size to be one of the strongest predictors of EHR adoption among hospitals. That result continues to appear in this data from 2012.

### **5.6.5 Limitations**

As in the physician analysis, this study would not have been able to provide a strong causal answer to the framing question, but it could potentially have illustrated a pattern of data consistent with one answer or another. In addition, the response rate was low (overall, around 25%), which could lead to significant self-selection or other bias. Additionally, survey questions regarding the adoption of EHRs are notoriously fragile – the particular form of the question have a significant impact on the response, making comparison of results across different survey instruments and data collection methods problematic. Finally, the overall number of responses was probably too low (about 150) to identify small effects.

### **5.6.6 Conclusion**

At least with respect to the impact of size on the likelihood of EHR adoption, the findings identified herein are not consistent with a significant effect.

## Chapter 6

# Have federal programs to promote the adoption of electronic health records changed the rate of adoption?

### 6.1 Background

The primary explicit goal for the programs providing financial incentives for the adoption of EHRs is to move the country toward universal EHR adoption. Given that EHR adoption had already been occurring, the implicit goal would seem to be increasing the rate of adoption. But, against what baseline? The adoption of most technology does not proceed along a generally linear, but rather follows an S-shaped curve. For several illustrative examples, figure 6.1 on page 83 shows the adoption curves for a number of common household products in the United States. While common household products may not be an appropriate comparison to EHRs, they are intended to be illustrative of the general trends and tendencies in technology adoption across a range of products.

While the adoption curves for these products track a range of shapes, most are typified by a low rate of adoption at first, followed by a significant uptick in adoption constituting the bulk of the range, followed by a long tail at a low rate toward universal adoption. This, then, would seem to support



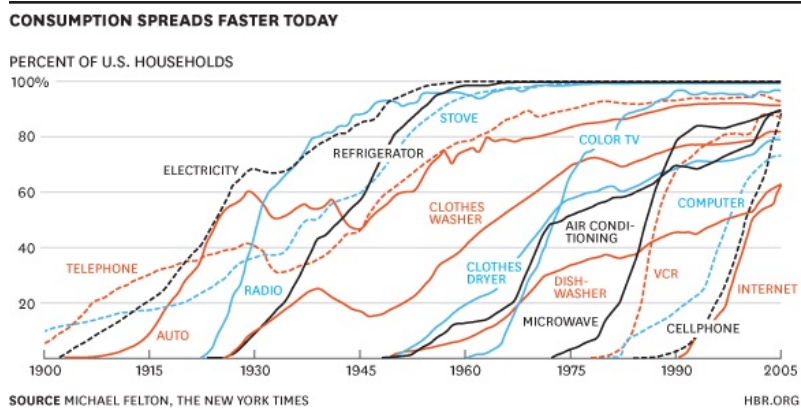


Figure 6.1: Adoption of Household Products in the U.S.

a preliminary claim that the appropriate baseline for EHR adoption against which to judge the effectiveness of programs attempting to increase these rates, is not a linear model, but rather an S-shaped curve of some variety. So, one of the first tasks will be to identify along which S-shaped curve EHR adoption was proceeding prior to the establishment of the EHR promotion programs, and then to identify along which S-shaped curve EHR adoption is now proceeding, and to compare the two. In the following analysis, I will do just that, starting with an attempted replication of some published results purporting to establish the baseline curve, followed by a critique of some of the methods used in the published work, wrapping up with some attempts to establish the prior and post curves using more robust methods.

### **6.1.1 Question**

How has the rate of EHR adoption changed since the establishment of the new federal EHR adoption programs?

### **6.1.2 Method**

Apply a Bass technology diffusion model to project future levels of adoption and compare the results to similar analyses using data from prior to the establishment of the federal EHR adoption programs.

### **6.1.3 Hypothesis**

Applying the Bass model to data that includes points in time after the establishment of the federal EHR adoption models will show a more rapid diffusion of EHRs into widespread use and a relative shift in the strength of external factors as opposed to internal factors in the diffusion. The formal hypothesis test will be whether measured adoption levels after the establishment of the federal programs differ significantly from projected adoption levels. The particular significance test will be identified through the analysis that follows.

### **6.1.4 Theory**

The explicit purpose of the federal EHR promotion programs is to create a more favorable environment for EHR adoption by lowering barriers to entry and increasing awareness. Smaller providers and those with fewer resources may be less likely to adopt EHRs. Therefore, providing additional fi-

nancial and technical resources through the federal EHR promotion programs could increase rates of adoption. If the actual rate of adoption is greater after the establishment of the federal EHR adoption programs, than predicted using only data from prior to the implementation of the federal EHR promotion programs, then it would suggest that one preliminary impact of the federal EHR adoption programs could be an increase in the rate of adoption.

### **6.1.5 Alternate hypothesis/theory**

The rate of adoption is similar or lower than estimated prior to the establishment of the federal EHR adoption programs. Increased pressure from external sources (i.e., the federal EHR adoption programs themselves) will induce resistance among providers, slowing a trend already underway in the absence of the programs.

## **6.2 Diffusion of Innovations**

The increasing adoption of a new technology within a community or market is often described as the diffusion of innovation. Originally popularized almost half a century ago, diffusion of innovation remains an active area of academic study and applied theory. In *Diffusion of Innovations*, first published in 1962, Everett Rogers, a rural sociologist, developed a broad theory that describes how technological innovations are adopted and spread across populations. Rogers reviewed over 500 studies of technology adoption and developed a theoretical framework describing many aspects of the diffusion of

innovation. Key aspects of Rogers theory include the elements of diffusion and the categories of adopters. According to the Rogers, the four primary factors that affect how a new innovation will spread throughout a population are innovation, communications channels, time, and a social system. The categories of adopters as defined by Rogers are innovators, early adopters, early majority, late majority, and laggards. Diffusion of Innovations remains influential and timely, with the fifth edition having been published in 2003. The theory laid out in Diffusion of Innovations was rendered into a quantified form by Frank Bass in the late 1960s. The Bass diffusion model was originally used, and continues frequently to be used to model the adoption of new consumer products. The original Bass article is the fourth most cited paper ever published in the journal Management Science. The EHR adoption literature includes the use of the Bass diffusion model to project future levels of EHR adoption. , The Bass diffusion model can be written as:

$$F(t) = (1 - e^{-(p+q)t}) / (1 + (\frac{q}{p})e^{-(p+q)t})$$

where F(t) represents the number of adoptions in time t, p is a coefficient representing internal pressures to adopt and q is a coefficient representing external pressures to adopt. This model has been applied in the EHR adoption literature by identifying EHR adoption rates at multiple points in time from multiple surveys and fitting the equation above to the several data points in order to estimate values of p and q.

Of additional interest, from the Bass equation, one can derive the “tipping point” by solving for the time, taking the derivative, and setting it equal

to zero. The resulting manipulation results in the following equation:

$$t^* = \ln(q/p)/(p + q)$$

### 6.3 Approach

In this section, I develop a data set of physician EHR adoption rates by year from the Texas physician survey discussed above, and fit the Bass diffusion curve to the data through the application of a linear optimization routine. In previous literature on EHR adoption using this approach, the data was taken from multiple surveys, giving rise to the possibility of significant imprecision due to instrumentation bias. After fitting the Bass model to the data prior to 2009, I develop confidence intervals, and compare the actual adoption levels after 2009 to the confidence intervals. Tracking the Bass exposition, changes in the relative values of  $p$  and  $q$  can be interpreted as a shift in the relative strength of internal versus external pressures to adopt the technology. The estimated  $p$  and  $q$  values can also be used to project levels of adoption and specific dates in the future and to compare these levels with projections based on  $p$  and  $q$  values derived from data collected only prior to the establishment of the federal EHR programs. Of particular interest in this section will be the methodological discussion. The structure of the Bass diffusion function would seem to make the shape of the curve fairly sensitive to relatively small variations in the underlying data, but little discussion of this point has been included in the published literature on EHR adoption that utilizes this approach. The methodological analysis will include sensitivity analysis to test

this possibility. The quantitative analysis will also include application of survival analysis techniques to the question of adoption levels at future points in time to serve as another test of the robustness of the findings from the Bass diffusion model.

#### **6.4 Past Projections of EHR Adoption Using the Bass Model**

There are two published papers applying the Bass model to the question of EHR adoption, both by Ford et al. Both papers utilized essentially the same method – identification of a series of annual EHR adoption levels by reviewing existing surveys and averaging in cases where multiple surveys were conducted in a given year. In the 2006 paper, only three annual data points were used, corresponding to 2001-2003. For the 2009 paper, seven annual data points were used, corresponding to 2001-2007. In both papers, multiple different surveys were used across years and within some years, multiple surveys were identified and generally combined through unweighted averages. Likewise, in both papers, key estimates included the Bass model parameters ( $p$  and  $q$ ), the projected adoption level in 2014, and the year of the tipping point. These several estimated values across multiple models, both those from the published literature and those developed herein, allow for comparison of the models and their relative predictive similarity or difference.

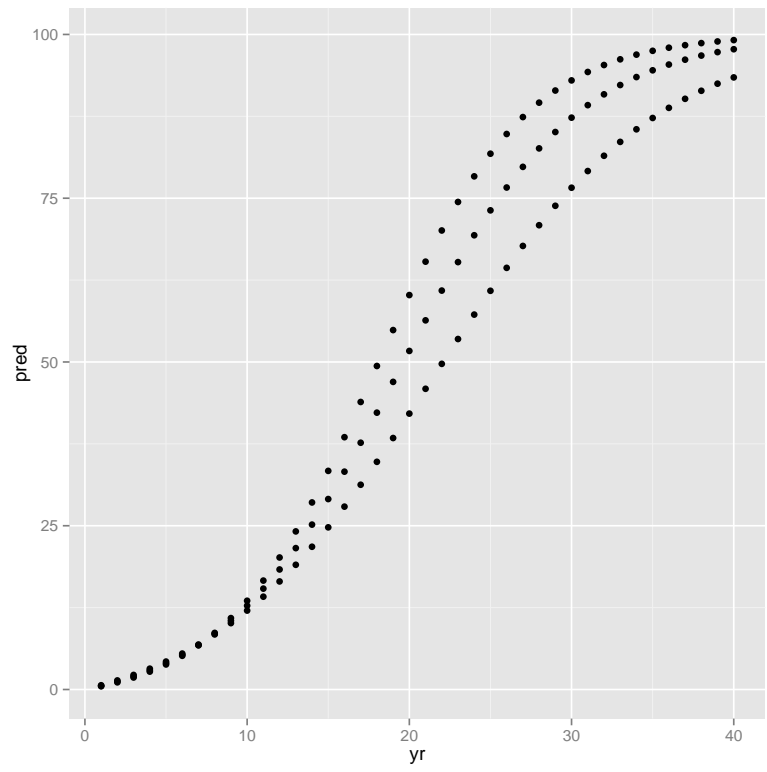


Figure 6.2: Ford (2006) - Best estimate, optimistic, and conservative

Figure 6.2 on page 89 shows the “best estimate” prediction along with “optimistic” and “conservative” projections based on the initial data points. (The x axis shows number of years since 1991.)

#### **6.4.1 Theoretical Problems with Past Approaches**

There are several theoretical and/or pragmatic problems with the approach taken in the Bass diffusion analyses of EHR adoption to date, primary of which is the potential incomparability of different survey results. EHR adoption rates based on provider surveys are notoriously dependent on the particular phrasing of questions regarding EHRs and adoption. Until recently, there were no standard terms for EHRs and it was difficult to know what providers were actually reporting when responding affirmatively to questions about EHR adoption. Therefore, using multiple different surveys to establish serial data points and/or average them for single years is fundamentally questionable.

#### **6.4.2 Statistical Problems with Past Studies**

Through correspondence with the authors of the previous studies, the underlying spreadsheets used to calculate the adoption levels reported in the papers and to develop the graphics included in the papers were obtained and analyzed.

While review of the papers provided much of the information necessary to recreate the published studies, analysis of the spreadsheets themselves



helped to identify some incongruities. As described by the authors, for both papers, they used the Microsoft Excel linear optimization tool with the summed differences between the estimated and actual adoption levels as the objective functions. The authors also indicate that an additional constraint was applied – that “the difference between the actual and estimated percentages of adopters for any year had to be less than 0.5% in absolute terms.” (Ford 2006; Ford 2009)

The authors were essentially attempting to fit the curve to the data, so why were additional constraints necessary? Analysis of the spreadsheets highlighted the issue. While the authors were clear to identify the objective function as summed differences between the estimated and actual adoption levels and, in fact, the objective function as calculated in the spreadsheets reflected this description, this is not a well-formed objective function. Merely summing the differences between multiple pairs of actual and estimated/calculated values leaves open the possibility that the differences could be offsetting with arbitrarily large positive and negative values that sum to near zero. As in the case of ordinary least squares regression, squaring the differences is necessary for the sum of these terms to represent the relative fit of the curve and serve as the objective function to be minimized. Using the spreadsheets provided by the authors themselves, upon removing the additional constraints, the models do not converge at all. In these models, the only thing that causes the estimated values to be near the calculated values is the set of additional constraints applied pairwise to the actuals and estimates.

### 6.4.3 Re-creation of Previous Studies

Re-structuring the objective function to calculate the sum of the absolute values of the differences between actual and estimated values leads to a better fit to the data and makes the additional constraints unnecessary.

	Residuals (Ford, 2006)	Residuals (Palmer, 2014)
2001	1.7668E-01	4.7824E-06
2002	-6.8104E-01	-9.9914E-01
2003	4.9695E-01	8.6609E-07

Table 6.1: Comparative Residuals

As shown in Table 6.1 on page 92, the differences between the “measured” values for the adoption levels in 2001-2003 and the calculated levels based on the Ford (2006) study are generally greater than the differences between the “measured” levels and the levels calculated in this analysis by minimizing the sum of the absolute values of the differences.

Scenarios	p	q	Tipping Point	2014 Adoption Percentage
Optimistic	0.0047	0.2001	2009	71.61%
Best Estimate	0.0054	0.1673	2011	61.93%
Conservative	0.0053	0.1544	2012	56.20%

Table 6.2: Estimates from Ford et al (2006)

Scenarios	p	q	Tipping Point	2014 Adoption Percentage
Optimistic	0.0044	0.2114	2009	74.43
Best Estimate	0.0051	0.1788	2010	65.25
Conservative	0.0060	0.1401	2013	53.50

Table 6.3: Estimates from Palmer (2014)

Comparing the results in Table 6.2 on page 92, which contains the estimates from Ford et al (2006), using the simple sum of differences and additional constraints to the results of this analysis as shown in Table 6.3 on page 92, small differences in the estimates are evident, but the relative relationships between the optimistic, best-estimate, and conservative estimates hold.

Another approach would be to estimate the model using the sum of the squares of the differences as in most regression analysis. The use of the sum of the squared differences in ordinary least squares and other types of regression analysis has a couple of key purposes and/or effects – it makes the resulting equations more easily soluble and/or transformable using well-established equalities and transformations, and it give greater weight to outliers. Using computational analysis renders analytic tractability superfluous while the Bass model’s non-linear parametric form may not benefit from overweighting outliers.

## **6.5 Sensitivity Analysis**

Another area within which the published work suffered from some unexamined assumptions was in their attempts at sensitivity analysis through the establishment of what might be called quasi-confidence interval. The 2006 paper included an effort to establish optimistic and conservative estimates associated with the diffusion curves in addition their “best estimate”. In order to develop the optimistic and conservative estimates, the authors added and

subtracted, respectively the standard errors and in one case, the standard deviation of the averaged values, from the annual point-estimates. So, for the optimistic estimate, the standard errors/deviation were added to each of the annual point estimates and the curve-fitting was calculated. For the conservative estimate the standard errors/deviation were subtracted from each of the annual point estimates and the curve-fitting was calculated. This gave rise to projected adoption curves that were higher and lower, respectively, than the “best estimate”. Unfortunately, this approach was econometrically flawed.

### **6.5.1 Extrema Permutation Analysis**

With a typical linear model, calculating expected dependent variable values based on the addition of standard errors to the coefficient estimates for the independent variables will consistently lead to a set of calculated dependent variable values higher than otherwise. Likewise the converse. However, with a non-linear parametric form, these assumptions do not hold. In the following analysis, curves are fitted to all permutations of the extrema, demonstrating that permutations other than “all high” and “all low” lead to greater variance in estimated adoption levels and tipping points. The reason for that this non-linear parametric form diverges in behavior from a linear form based on the variation in the input data is that the curvature at all points of the function is determined by the available data. Thus, if the curve is fitted to just a few points toward the low end of the function, and those points describe a curve that is particularly concave-up, then the subsequent curve-fitting may result

in an equation for which the upper part of the function is particularly concave-down, which may imply higher adoption rates at earlier points in time. A set of data points mixing high and low estimates for each point in time would lead to a set of points describing a more concave-up curve.

Year	Measured Adoption Level	Std Err/Dev	Optimistic	Conservative
2001	12.80	0.75	13.55	12.05
2002	14.40	1.27	15.67	13.13
2003	18.32	1.83	20.15	16.50

Table 6.4: Adoption Levels and Standard Errors/Deviations

Table 6.4 on page 95 shows the measurements that were used by Ford et al (2006) for the Bass optimization and the sensitivity analysis thereof. For the “optimistic” estimate, the standard errors/deviation were each added to the corresponding year’s measured adoption level and for the “conservative” estimate, they were each subtracted from the adoption levels.

2001	2002	2003	p	q	2014 Level	Tipping Point
13.55	15.67	20.15	0.0044	0.2114	74.43	2009
13.55	15.67	16.50	0.0122	0.0359	33.92	2013
13.55	13.13	20.15	0.011	0.0567	37.74	2015*
13.55	13.13	16.50	0.0124*	0.033+	33.43+	2013
12.05	15.67	20.15	0.0023+	0.2909*	86.72*	2008+
12.05	15.67	16.50	0.006	0.1401	53.5	2013
12.05	13.13	20.15	0.0023+	0.2909*	86.72*	2008+
12.05	13.13	16.50	0.006	0.1401	53.5	2013

Table 6.5: Extrema Permutations

Comparing the results of the extrema permutation analysis as shown in table 6.5 on page 95 to the results from the sensitivity analysis conducted

using only the high and only the low extrema as shown in table 6.3 on page 92 shows that there are, in fact, permutations of the extrema other than “all high” and “all low” that are greater and lesser, respectively, representing a range that could more appropriately be described as “optimistic” and “conservative” estimates. (In this table, the largest values in each column are indicated by a \* and the smallest by a +.)

### **6.5.2 Monte Carlo Modeling to Test Extrema Permutation Assumptions**

Although it appears likely that permutations of the extrema would lead to the identification of the greatest variance in the predicted future adoption levels and tipping points, a Monte Carlo modeling approach will provide greater confidence that these conclusions are robust. In the following analysis, the future adoption levels and tipping points are calculated based on annual values of the adoption levels for the three years of interest that are randomly selected from the range established by the standard errors/deviations. In this analysis, the results of cycles of 10, 100, 1,000, and 10,000 iterations are shown. As predicted, the calculated future adoption levels and range of tipping points appear to converge to the values estimated through the extrema permutation analysis.

In order to assess whether the extrema permutation approach provided the maximum range of possible estimates under the range of scenarios (“optimistic” and “conservative”), Monte Carlo runs of 10-1000 iterations for each

of the variables of interest, each for both the “optimistic” and “conservative” estimates. Figure 6.3 on page 98 and 6.4 on page 99 show two respective results of these Monte Carlo runs. In each figure, the horizontal line represents the maximum or minimum of the calculations from the extrema permutation analysis and the scatterplot is overlaid by a loess fitted curve. In a few of the cases, there are projected values of the variables outside the bound established by the extrema permutation analysis, but in all cases, these variances are small. Therefore, the hypothesis that the extrema permutation analysis would yield appropriate “optimistic” and “conservative” estimates based on the standard errors/deviations of the original data points is supported.

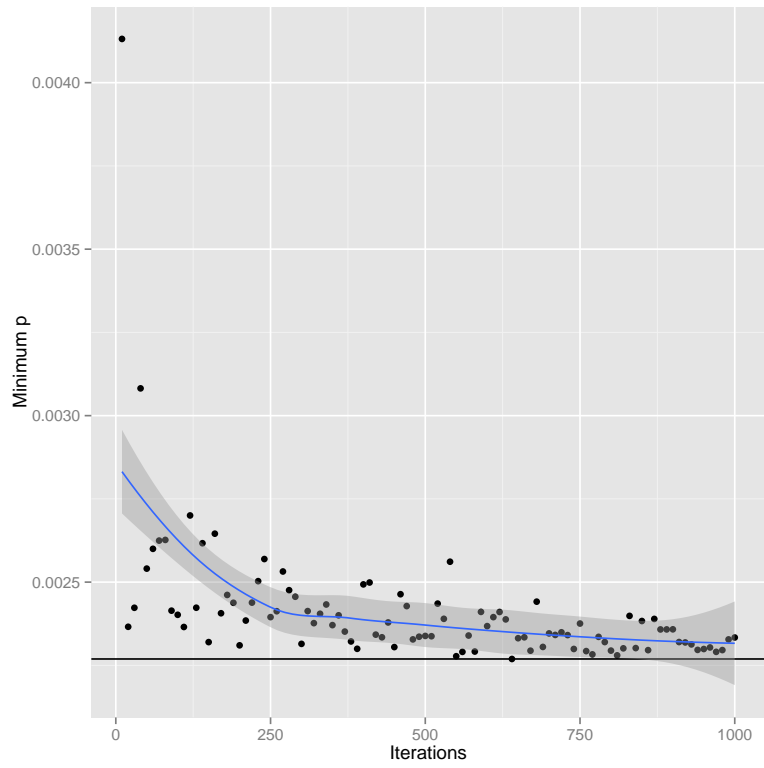


Figure 6.3: Monte Carlo Results for Minimum p



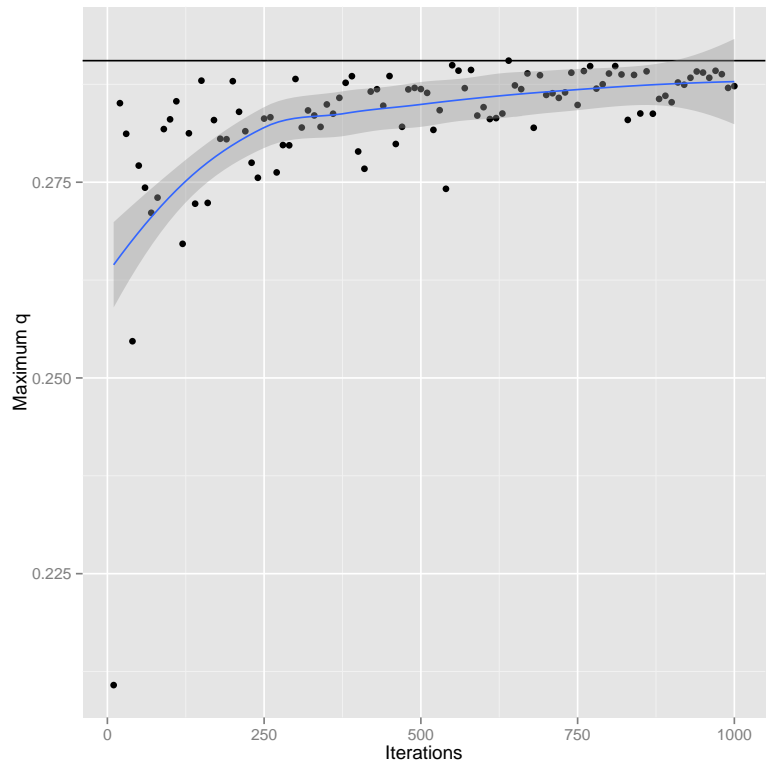


Figure 6.4: Monte Carlo Results for Maximum  $q$

## 6.6 The Bass Model: a Reprise

And so we return to the Bass model. The initial versions of the Bass model analyzed herein involved the aggregation of results from multiple, different surveys. This approach had some potential theoretical problems due to the possibility of instrumentation bias – that the different surveys might have been measuring different things. Consideration of an alternative approach – survival analysis – did not prove as fruitful as initially assumed, given the realities of the data. Fortunately, the practitioner survey data itself can be re-worked and re-structured in way that allows for the application of the Bass diffusion model, without the issues resulting from multiple, different surveys. The 2012 practitioner survey, in particular, is amenable to this sort of processing. Since the survey asked practitioners when they adopted an EHR, a histogram of the resulting answers can be constructed, which represents the number of practitioners who adopted EHRs in each year. This data can then be processed into a cumulative distribution which can be used as the input data for the Bass model.

Figure 6.5 on page 102 shows the number of EHRs adopted in each year and Figure 6.6 on page 103 shows the cumulative adoption percentage. It should be noted that the survey was conducted between June 21, 2012 and October 12, 2012, so the number of physicians reporting full or partial adoption in 2012 probably understates the proportion of the sample that actually did achieve full or partial EHR adoption during 2012.

In order to account for this timing, a portion of the physicians reporting

that they intended to adopt within the next year are randomly selected and allocated to the full and/or partial adoption category for 2012 (both of which were coded to the same value for analysis). The apportionment approach is to select a fraction of the physicians reporting that they intend to adopt within the next year that is proportional to the fraction of the year remaining from the mid-point of the survey period. It is important to note that the subsequent projections of adoption rate and development of confidence intervals around that projection only use data on adoption through 2009 in order for the projection to be a reasonable counterfactual to observed adoption rates after the establishment of the EHR promotion programs.

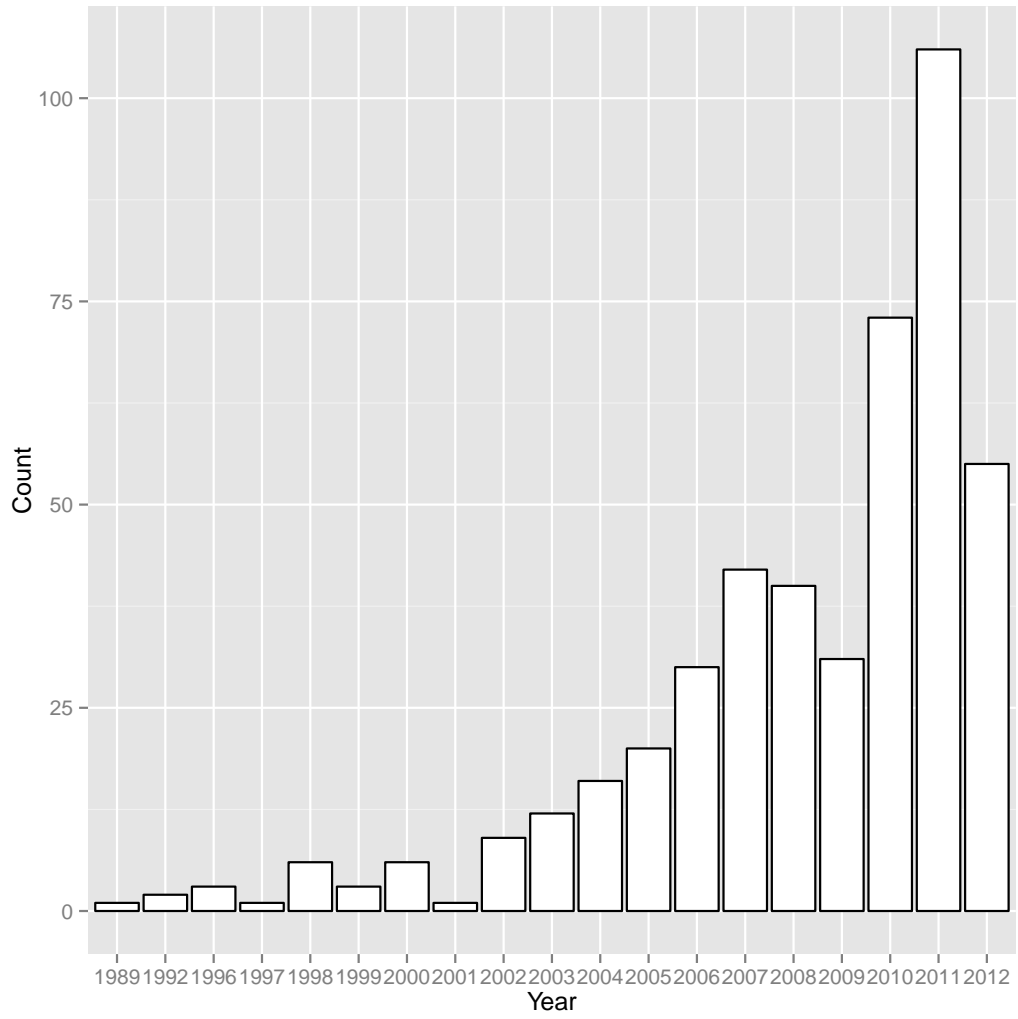


Figure 6.5: Adopting Physicians by Year

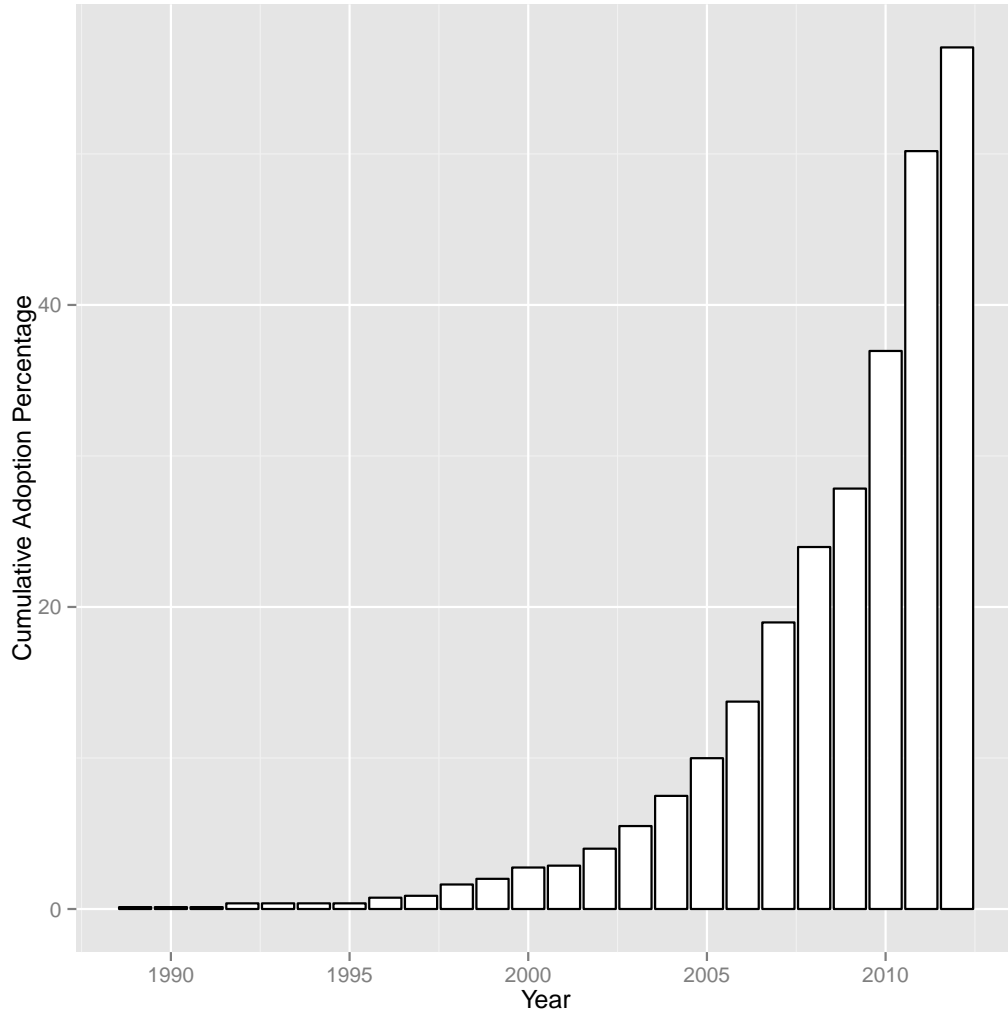


Figure 6.6: Cumulative Distribution

	p	q	2014 Level	Tipping Point
Pre	0.00	0.33	69.27	2012.00
Post	0.00	0.35	72.74	2011.00

Table 6.6: Estimated Parameter Values

Table 6.6 on page 104 shows the p and q values calculated based on the cumulative percentage curves for the time period through 2009 (“Pre”) and then including 2010 and 2011 (“Post”). In this case, “Pre” and “Post” refer to the time period prior to the implementation of the EHR incentive and promotion programs and “Post” includes the years after their establishment. The values in the “Pre” and “Post” series are virtually identical, suggesting that the data fall on virtually the same adoption curves. Ex ante, it would have been reasonable to hypothesize that the “Post” adoption curve would project a higher level of adoption. These findings do not support that hypothesis.

## 6.7 Building a Better Confidence Interval

Although Ford et al did not describe their optimistic and conservative projections of EHR adoption rates as confidence intervals, given their mathematical basis in the standard errors and deviations of the underlying data, as well as their characterization within the two Ford papers and the absence of any other effort to generate confidence intervals, they are serving that role. However, as described in the preceding discussion, there were numerous problems, both theoretical and practical, with the approach taken by Ford et al. In this section, I will develop confidence intervals using bootstrapping, following

the approach offered by Davison and Hinckley (1997).

The cumulative adoption percentage as shown in figure 6.6 on page 103 represents the cumulative rate of adoption of EHRs by physicians, as measured by year. This graphic rendering of the cumulative adoption percentage aligns with the typical graphic depiction of technology adoption as predicted by the Bass diffusion model, with adoption percentage plotted on the y-axis as a function time measured on the x-axis. Thus, if the adoption rate of EHRs follows the bass diffusion model, then this cumulative adoption rate, projected forward using the Bass formalism, should asymptotically approach 100%.

While the underlying survey data was collected across just a few months in 2012, the numbers of EHR adopting physicians by year and the cumulative adoption percentage across time can be derived from the underlying data due to the inclusion of a question regarding date of adoption. For the development of the confidence intervals, I use a bootstrapping approach through which new samples are generated by selecting samples from the underlying data of a size equal to the underlying sample size, with replacement, then re-calculating the cumulative adoption percentage levels, and using these values for a re-estimation of the parameters in the Bass diffusion equation. For each new set of parameters, the Bass diffusion equation is used to calculate projected adoption levels at each year through 2014. These values are then stored, and the process is repeated 999 more times. Finally, for each year, the projected adoption levels at the 5% and 95% levels are identified (i.e., for the set of 1,000 projections for each year, the 51st and 950th values, respectively, from

the sorted list were identified). The set of projected adoption levels at the 95% level defines the upper confidence interval, and the set of projected adoption levels at the 5% level traces the lower confidence interval. The new confidence intervals can be seen in figure 6.7 on page 106 (dashed blue lines) along with the projected adoption level (solid blue line) and actual adoption levels (black points).

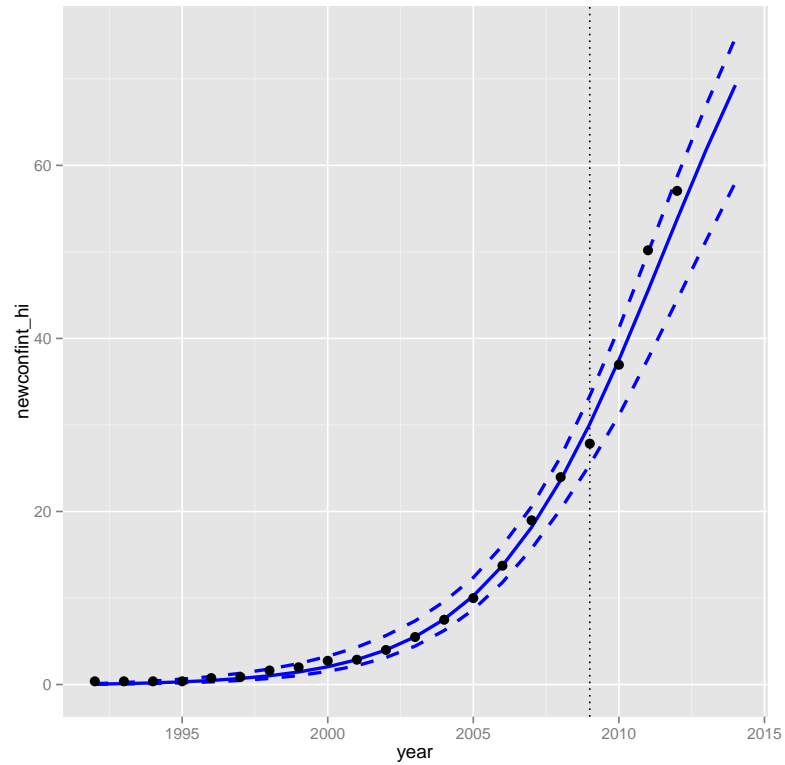


Figure 6.7: Basic Confidence Interval Plots with Bootstrapping



## 6.8 Hypothesis Test

Since the underlying question motivating this whole analysis is whether the federal programs established to promote the adoption of EHRs have had an effect, one useful element in trying to answer the question is the establishment of a believable counterfactual. That is, what would have happened if the programs had not been established? Using the cumulative EHR adoption levels as gathered from the 2012 physician survey, it is possible to estimate the Bass diffusion model model for times through the present and beyond, using only data from before the establishment of the federal EHR promotion programs. The core assumption in taking this logical step is that, in the absence of the EHR promotion programs, the rate of EHR adoption would have tracked the Bass diffusion curve, as estimated based on the cumulative levels of adoption prior to the establishment of the programs. Therefore, the formal hypothesis test is whether the actual measured adoption levels for years after the establishment of the federal EHR promotion programs are within or outside the confidence intervals calculated based on data only from before the establishment of the federal EHR promotion programs. visually reviewing the plot with confidence intervals and actual measurements as shown in figure 6.7 on page 106, it appears that two of the year measures are very close to the boundary of the confidence interval. Review of the actual underlying measures and estimates, as shown in table F.3 on page 195 shows that one of the year measures is, in fact, outside of the 95% confidence interval. This would lead us to reject the null hypothesis and, subject to the other assumptions of the

model, accept the alternate hypothesis that the programs had an effect, at least as measured in that one year.

Year	Conf. Int., Low	Measured Adoption	Conf. Int., High
2010	31.11	36.95	41.18
2011	37.66	50.19	49.89
2012	44.45	57.05	58.69

Table 6.7: Actual Adoption Levels and Confidence Intervals

## 6.9 Discussion

The analysis in this chapter attempted to accomplish several goals – analysis and replication of previous attempts to project future adoption rates using the Bass diffusion model, consideration of survival analysis as an alternate analytic framework, and extension of the Bass model approach. Previous attempts to apply the Bass diffusion model to EHR adoption suffered from theoretical and statistical problems. In particular, the use of data from multiple sources could introduce unpredictable measurement errors, and the particular approach to the identification of confidence intervals was problematic. A more appropriate calculation of confidence intervals yields a substantially broader range than appears in previous (published) efforts.

Using the 2012 practitioner survey, it was possible to construct a data series that was amenable to optimization using the Bass diffusion model. Such analysis, comparing the data prior to the implementation of the several programs to promote the adoption of EHRs, and the data series including dates

after the initiation of the programs, showed little difference in the adoption curve, a finding consistent with minimal impact on adoption rates.

Survival analysis showed initial promise in describing the adoption pattern, but, given the lack of time-based covariates in the data, did not show any marginal benefit as a tool for projecting future adoption levels, although applied to a different data set it might be productive. The preliminary steps associated with the survival analysis approach are included in Appendix E for reference.

Although the federal EHR promotion programs were all authorized at a particular point in time in early 2009, their programmatic and fiscal effects emerged over several years. Therefore, it is relevant to consider whether 2009 should be taken as the point-in-time against which to judge whether they had an effect. Additional quantitative analysis considering later years as the effect time are included in Appendix F.

## **6.10 Limitations**

In addition to having the limitations of previous analysis contained in this dissertation due to instrumentation challenges associated with attempts to measure EHR adoption by survey, generalizing from this particular analysis also is also limited by the research design. Although the use of data from a single survey instrument should lower the risk of instrumentation bias, the lack of a control group or other mechanism to eliminate other possible causes means that, at best, we can only assert that the findings are consistent with

no effect.

## **6.11 Conclusion**

Past analysis has predicted future physician EHR adoption rates by fitting observed EHR adoption rates from multiple surveys to the Bass diffusion of innovation model. Although these studies had statistical and theoretical issues, they predicted EHR adoption rates that are consistent with those that have been observed since their publication. More importantly, however, predictions of future adoption rates based on data series including only years prior to the federal EHR incentive programs and predictions of future adoption rates based on data series including years after the establishment of the federal EHR incentive programs yield similar values, suggesting that the federal EHR incentive programs have not had an impact on the adoption of EHRs by physicians.

## Chapter 7

# Have federal programs to promote the adoption of electronic health records changed the level of adoption?

### 7.1 Overview

There are three main federal programs to promote the adoption and "meaningful use" of certified EHRs by physicians. The Medicare EHR incentive program will pay each qualifying doctor about \$45,000 total over 5 years for the "meaningful use" of a certified EHR. The Medicaid EHR incentive program will pay each qualifying doctor about \$64,000 total over 6 years for the adoption and "meaningful use" of a certified EHR. Finally, the HIT Regional Extension Center (REC) program established a national network of organizations to assist physicians in small primary care practices. The federal programs to select, adopt, and become meaningful users of certified EHRs. To qualify for technical assistance from a REC, a physician group had to focus on the delivery primary care, and have 10 or fewer physicians.

The purpose of this chapter is to try to directly assess the impact of the federal programs to promote the adoption of EHRs by taking advantage of some of the programmatic features in the analysis. In particular, there are

aspects of the HIT Regional Extension program for physicians and the Medicaid EHR incentive program for hospitals that lend themselves to a regression discontinuity design.

## 7.2 Regression Discontinuity Design

The regression discontinuity (RD) design was described and popularized by Shaddish and Campbell [cite]. The RD design is generally characterized by an exogenous cut-off value, that splits a population into two groups. As long as the cut-off value is not tied, a priori, to any particular characteristic of the underlying phenomenon, then its effect can be considered similar to random assignment

In general, we would not have any reason to assume that observations just on one side of the threshold vary in any meaningful way from observations just on the other side of the threshold. With only the observations on one side of the threshold receiving the treatment, the observations on the other side of the threshold can be considered similar to a 'true' control group in that they are similar in all meaningful ways to the units receiving the intervention, and they are selected by a process that does not systematically bias the characteristics of the units in the group. Thus, RD design is generally considered to be a strong research design for attributing causality since it would be reasonable to assume that units of analysis on either side of the cutoff, particularly near to the cutoff, should be very similar, making appearance on one or the other side of the cutoff serve a role similar to random assignment in a classic randomized

controlled experiment.

The RD design, under the right circumstances, can provide the basis for causal inference as well as a randomized, controlled experiment. One of the key circumstances is rigid adherence to the cut-off. The units of analysis cannot have precise control over which side of the cut-off they are on. Even though they may be able to influence their measurement on the running variable (as the variable along which the cut-off applies is often referred), as long as they are not able, precisely, to determine which side of the cut-off they are on, the underlying assumptions supporting causal reasoning from the RD design still hold.

In this case, the running variable is physician group size and cut-off is at 10. While physicians can certainly control the size group in which they practice, it seems unlikely that physicians would be changing their practice structures (and sizes) just to be eligible for this program. On the other hand, physicians may maintain individual or small-group practices, while belonging to larger physician organizations for purchasing or contract negotiating purposes. As such, even though some physicians may enjoy the resources that come along with a larger group, they may technically retain their small practice status and be eligible for the REC program. I will return to this question in the subsequent discussion with some specific analysis regarding the likelihood of sorting around the cutoff.

### **7.3 Methodological Note Regarding Linear Models**

The use of linear regressions in this chapter is worth noting because it diverges from the use of logistic regression in a previous chapter for estimating the contribution of a number of different factors to the adoption of EHRs where the unit of analysis was the individual physician or hospital. With such data, logistic regression is generally preferred because the estimated values are bounded by zero and one, as would be expected when the range of answers is only zero and one.

A common concern regarding the use of linear regression for situations where the the dependent variable is a binary variable or a percentage is the possibility that the model might estimate values less than zero or greater than one. While this concern may be justifiable on intellectually aesthetic grounds, in cases such as the RD design where the particular data of interest is far away from the portions of the data that might generate estimates less than zero or greater than one, the use of linear regression would appear justified based on its simpler interpretation. In this case, the use of linear regression rather than logistic regression to estimate the likelihood of an outcome with only binary options implies that the underlying model is a linear probability model.

### **7.4 Impact of Regional Extension Centers on Physician EHR Adoption**

The HIT Regional Extension Center (REC) program was established to provide technical assistance to primary care physicians in small group prac-



tice. The technical assistance was intended to include EHR product selection, implementation, workflow re-engineering, and meaningful use. Texas had four recognized and funded RECs that collectively received about \$36 million over the four years of their operation, with the explicit goal of assisting about 6,000 primary care physicians in Texas to become meaningful users of their EHRs.

For the purposes of this analysis, the two important programmatic characteristics are that the RECs were generally only funded to assist physicians in small group primary care (10 or fewer clinicians) . So, with respect to the RD design, the relevant values and thresholds are primary care specialty and practice size of 10 or fewer. There would appear to be no reason, 'a priori', to believe that there is anything special or particular about the program threshold of 10, making it a good candidate for an RD design breakpoint.

#### **7.4.1 Data**

The data used in this chapter is the same 2012 physician survey data, collected by HHSC, used in previous chapters. After limiting the practitioner survey to physicians only, eliminating rows with incomplete data, and merging with the county data file, 813 records remained. The practitioner survey data is merged with county data, allowing for the inclusion of socio-economic and other demographic/geographic characteristics, without losing any data. Finally, since the question of interest in this section is whether the REC program had an impact on the adoption of EHRs by primary care physicians, and I am trying to take advantage of a regression discontinuity design, physi-

cians who had already adopted EHRs prior to the establishment of the REC program and specialists are removed, leaving 249 records.

Although typically an RD design is characterized by pre and post intervention measurements, this data includes a variable capturing the date of EHR adoption, which allows us to identify the pre-intervention EHR adoption status. The post-intervention EHR adoption status is then identified through the simple question regarding current EHR adoption status.

Also of note, since the strength of the RD design leverages the random sorting near the cutoff, the analysis in practice tends to focus on data in the vicinity of the origin. As will be seen during the discussion of the results, the data is heavily weighted toward the origin. There are a few outliers in the opposite direction that are removed from subsequent analysis due to being so far removed from the cutoff. Specifically, recalling that the running variable is physician group size and the remaining data set contains 249 records, 245 of the records represent physicians practicing in groups of 20 or fewer. Of the four truncated records, one represented a physician practicing in a practice of 35, one a practice of 100, and two in practices of 1,000.

In addition to being truncated due to distance from the cutoff, and concerns about the effect they would have on polynomial estimates of data near the cutoff, it is notable that several of these numbers are so round. It seems unlikely that the physicians in the group practicing in the largest groups were practicing in groups of exactly 100 and 1000. Therefore, it is likely that these and perhaps others of the reported practice sizes were inexact estimates

reported by the physicians.

#### **7.4.2 Theory**

One challenge that has often been cited in physician EHR adoption is the lack of technical knowledge, particularly in smaller groups, and the lack of resources to purchase the required expertise. Providing a cadre of experts in EHR adoption and implementation at no cost to the physicians could mitigate these challenges as was done through the Health IT Regional Extension Center (REC) program.

#### **7.4.3 Hypothesis**

I would anticipate that physicians eligible for technical assistance from the REC program would have adopted EHRs at a rate greater than those who did not have access to the assistance. Within the context of the RD design, if the REC program increases the level of adoption among eligible physicians, then I would expect to find a statistically significant jump in the value of the dependent variable at the eligibility threshold (i.e., physician group of 10 or fewer.)

#### 7.4.4 Results

In my analysis and presentation of results, I roughly follow the approach described by Lee and Lemieux (2009). Both local linear regression with different bandwidths and multiple polynomial forms are tested. A strong result will show statistical significance and similar effect size across multiple bandwidths and parametric forms.

##### 7.4.4.1 Assessing Parametric Form

An important consideration when applying the RD design is the choice of parametric form. Typical approaches include the use of local linear regression with differing choices of bandwidths and series estimation whereby polynomials of higher order are fit to the data. Figure 7.1 on page 119 shows the underlying data, plotting EHR adoption against practice size, a loess-fitted curve with confidence interval, and the program eligibility cut-off (physician group size of 10 or fewer).

Since EHR adoption is a binary variable and practice size as a variable only takes integer values, the basic x-y plot of these variables results in a large amount of overplotting, making the underlying density of the data difficult to discern. In this plot, jitter is introduced, shifting the individual data points by a random amount within a small x-y window, and point opacity is reduced, rendering individual points somewhat transparent and clusters of points darker.

There are several important things to note about this figure. First, the

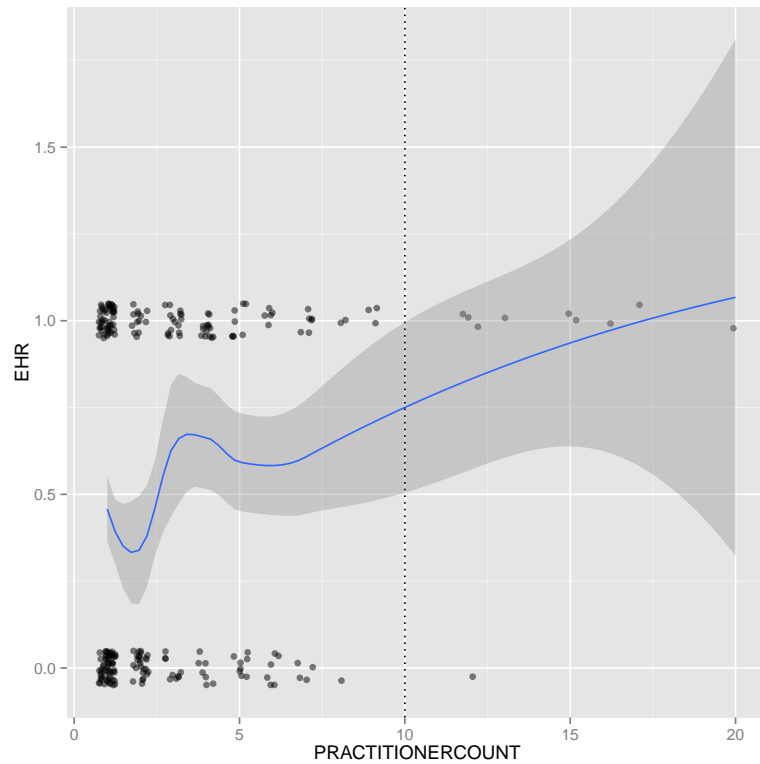


Figure 7.1: Adoption vs Practice Size - Data, Fitted Curve, and Cutoff

purpose of graphing a non-parametric fitted curve to the data when using the RD design is to get a sense of what the underlying parametric form might be. With three points where the derivative of the fitted curve appear to equal zero, it would be reasonable to assume that fourth-degree polynomial would be a good fit. The subsequent analysis includes polynomial estimates up to the quartic form.

Second, the data density approaching and beyond the cutoff becomes very low. This low data density has two main implications – the confidence

interval blows up as the fitted curve approaches and goes beyond the cutoff, and data density around the cutoff is low. The latter of these points is the more salient for an RD design. Although the treatment effect as measured at the cutoff can be interpreted as a weighted treatment effect across the full range of the data, in practice, treatment effects estimated based only on data near the cutoff seem to be preferred. Thus, if a treatment is estimated primarily or only using data near the cutoff, with this low level of data density, the standard errors will be very high.

Third, the data is heavily skewed toward the origin. For any local regressions or polynomial estimates that include the high density of data near the origin, it appears likely that the heavy data density near the origin may bias the estimate of the treatment effect at the cutoff.

#### **7.4.4.2 Assessing the Possibility of Sorting near the Cutoff**

The next step in the analysis and presentation of the results of a research question structured using an RD design is to assess whether there appears to be differences in data density immediately on either side of the cutoff that would suggest precise sorting by the units of analysis. Both visual and analytic approaches are available for this test. First, as shown in figure 7.2 on page 121, the data density is plotted against the running variable. (As in the previous figure, the cutoff is indicated with the dotted line.) Also shown is a loess-smoothed curve tracking the density.

There does appear to be a higher data density just below the cutoff,

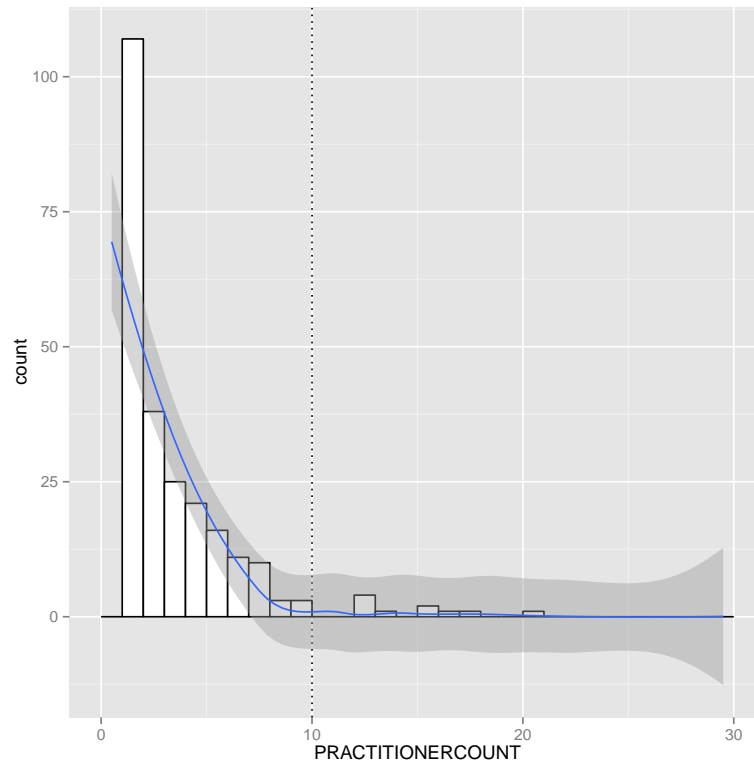


Figure 7.2: Practice Size Distribution, Fitted Curve, and Cutoff

which might suggest physicians somehow gaming their practice sizes in order to qualify for technical assistance from the RECs. A more formal approach would be preferable, but the density comparison calculation suggested by McCrary (2008)<sup>1</sup> fails due to lack of adequate data density around the cutoff. The observation of possible sorting around the cutoff will become part of the further discussion regarding limitations of this study.

---

<sup>1</sup>McCrary 2008.

### 7.4.4.3 Optimal Bandwidth Calculation

The next step in the analysis and presentation is to determine if there is an optimal bandwidth for the local linear regression approach. Following the approach suggested by Imbens and Kalyanaraman (2011)<sup>2</sup>, the optimal bandwidth is calculated as 5.0604.

### 7.4.4.4 Graphing Different Parametric Forms

The next step in the analysis and presentation is the graphical display of estimates based on several different bandwidths and parametric forms. Figure G.1 on page 199 shows the graph of linear models with bandwidths 5 and 10 and polynomial estimates for quadratic, cubic, and quartic functions based on data across the whole of the remaining data space. Figure G.2 on page 200 is a magnified version of the region of figure G.1 outlined by the dotted rectangle.

---

<sup>2</sup>Imbens and Kalyanaraman 2011.



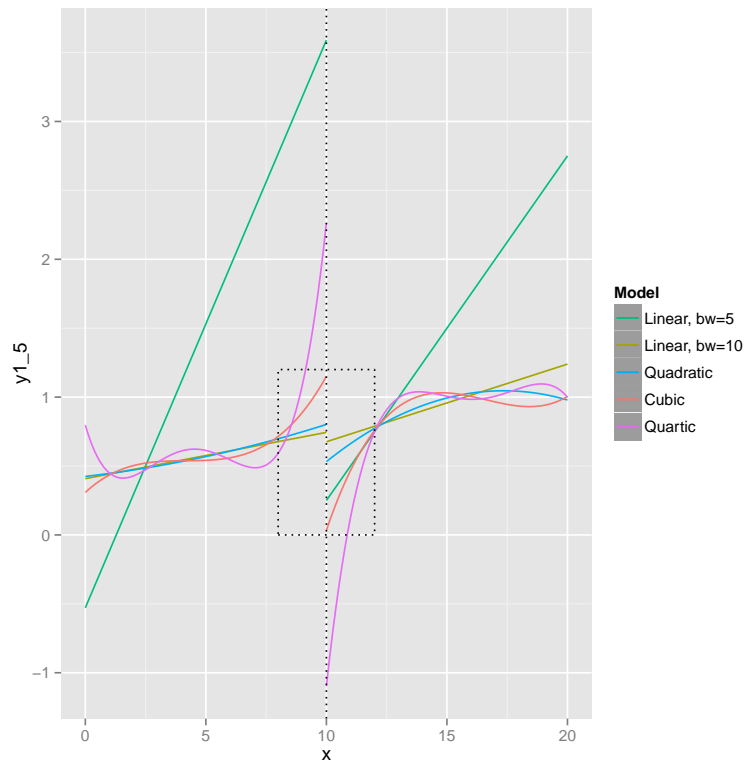


Figure 7.3: REC RD Design Table - Multiple Models

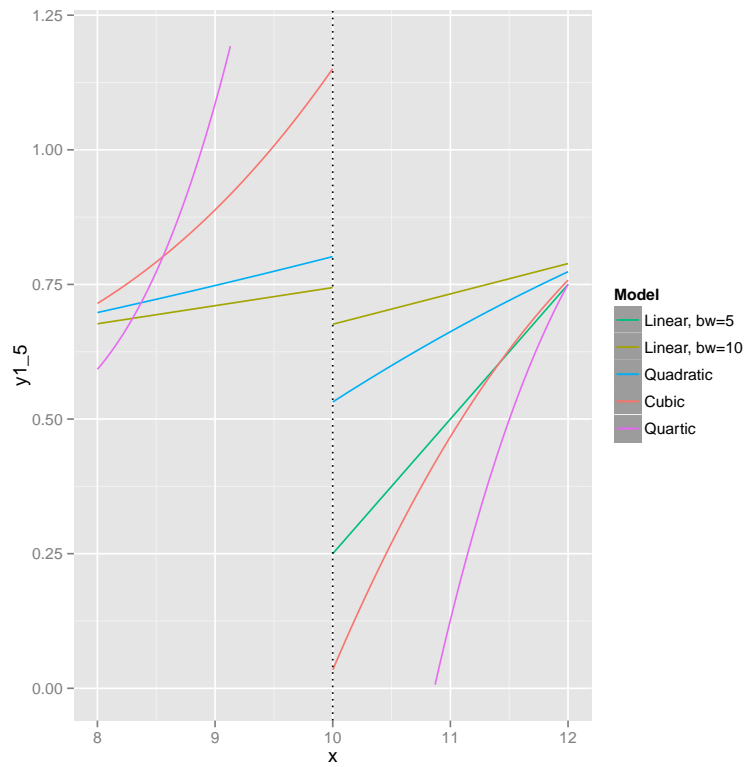


Figure 7.4: REC RD design Table - Multiple Models, Cut-off zoom

Common practice for the presentations of results from analyses using the RD design is to estimate local linear regression models using multiple different bandwidths, often multiples of the optimal bandwidth, in order to test the robustness of the results. It is also common to test other parametric forms against the full range of the data based on similar logic – that consistent effect estimates and statistical significance across a wide range of specifications supports greater confidence in the estimate.

Figures G.1 and G.2 contain too many different estimates to include confidence intervals for all of them in a way that could be visually discernible so the effect sizes and relative significance levels are summarized in table G.1 on page 201.

	Estimate	Pr(> t )
Local linear, bandwidth=5	1.7213	0.7978
Local linear, bandwidth=10	0.2957	0.8091
Quadratic	2.2449	0.6852
Cubic	18.0058	0.7764
Quartic	129.5635	0.8381

Table 7.1: REC RDD Effect Sizes at Cut-off

#### **7.4.5 Discussion**

The Health Information Technology Regional Extension Center (REC) program was established to help primary care physicians in small practices adopt and become meaningful users of certified electronic health record (EHR) systems. Using data on physician EHR adoption in Texas, I was able to identify a set of physicians who had not adopted an EHR at the beginning of the REC program, and then analyze the subsequent pattern of adoption, using eligibility for assistance from the REC program as the cutoff value on the running variable. The results identified herein are inconclusive. Although the effect sizes as estimated across a number of different models are all in the same direction, which is also the direction that one would expect if the program were successful, none of them are statistically significant.

In addition, Appendix G includes additional analysis in which covariates are added to the RD design models discussed above. Finally, Appendix H includes additional analysis in which additional bandwidths are tested with all of the parametric forms discussed above.

#### **7.4.6 Limitations**

Some of the limitations of the physician RD design analysis include the self-reported nature of the data, the limited sample size, the low response survey response rate, and recognized imprecision in responses to questions regarding EHR adoption.

## 7.5 Hospital Medicaid EHR Incentive Program Impact

In this section, I analyze the impact of the Medicaid component of the EHR financial incentive program on hospital EHR adoption. Federal law provides for a system of financial incentives for hospitals that make meaningful use of certified electronic health records. As in the case of the corresponding program for physicians, there are versions of this program authorized for administration by both the federally-administered Medicare program and the state-administered Medicaid program. Unlike the case of the physician programs, hospitals are eligible to receive financial incentives from both Medicare and Medicaid. The difference is justifiable from a policy perspective because the amount of funding for which each hospital may be eligible is determined based on formulae that consider the relative volume of Medicare and Medicaid patients, whereas the physician incentives are paid on a simple rate schedule (that does differ between Medicare and Medicaid).

The Medicare EHR incentive program requires that the hospital be a Medicare provider, with no additional volume-based eligibility criteria. The Medicaid EHR incentive program, in contrast, requires that hospitals have a Medicaid volume of at least 10%. This program feature can be exploited by an RD design since there is no obvious theoretical reason to believe that hospitals immediately on either side of that cutoff value vary in any systematic way from each other.

### **7.5.1 Data**

The data used is the same as in the previous logit analysis – hospital survey data collected by HHSC in 2012, joined to county-level geographic and socio-economic information. As in the case of the physician RD design analysis, the data is subset to include only those records corresponding to hospitals that had not yet adopted EHRs prior to 2009. The original data set had 177 records. After eliminating several records due to incomplete data, merging the remaining data with county and zipcode level data, and removing records that corresponded to hospitals that had already adopted an EHR prior to 2009, there were 91 remaining records, almost exactly half of which reported that they had adopted an EHR in the 2012 survey.

### **7.5.2 Theory**

As in the case of physicians, the conventional wisdom has held that implementation and transition costs were among the most significant impediments to hospital EHR adoption.

### **7.5.3 Hypothesis**

Provision of additional financial assistance should increase the marginal likelihood of hospital EHR adoption by hospitals. Other factors being equal (or close to equal), if cost is, in fact, an impediment to hospital EHR adoption, then hospitals receiving the additional amount of funds from Medicaid should adopt at a higher rate. As estimated using the RD design, hospitals close to

the cutoff level of 10% Medicaid volume on either side would not be expected to differ in any meaningful way from each other, establishing the pre-conditions for a quasi-experimental design that is as good as a randomized controlled experiment for supporting claims of causal inference. Thus, a result showing a statistically significant effect size at the cutoff, with a value of similar scale and the same direction across multiple model specifications, should be interpretable as a causal effect.

#### **7.5.4 Results**

The presentation of results for the hospital analysis closely follows the presentation of results for the physician analysis, with only minor differences due to the cutoff occurring against a different running variable. As such, the narrative will be slightly less verbose, but will follow the same general contours.

##### **7.5.4.1 Assessing Parametric Form**

As in the physician analysis above, here it is necessary to assess different potential parametric forms. As in the case of the physician analysis above, models estimated will include local linear regression with differing choices of bandwidths and series estimation whereby polynomials of higher order are fit to the data. Figure 7.5 on page 130 shows the underlying data, plotting EHR adoption against Medicaid volume, a loess-fitted curve with confidence interval, and the program eligibility cut-off (Medicaid volume of 10% or greater).

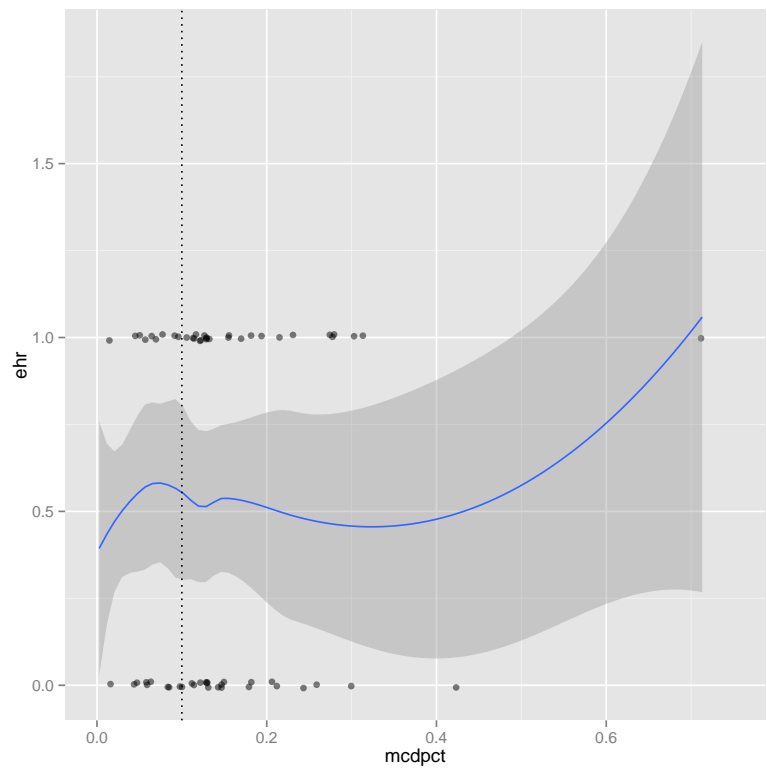


Figure 7.5: Hospital Medicaid % and EHR Adoption - Data, Fitted Curve, and Cutoff



Similar to the case of the physician data, since EHR adoption is a binary variable and Medicaid volume as a variable is relatively tightly clustered, the basic x-y plot of these variables results in a large amount of overplotting, making the underlying density of the data difficult to discern. In this plot, jitter is introduced, shifting the individual data points by a random amount within a small x-y window, and point opacity is reduced, rendering individual points somewhat transparent and clusters of points darker.

As in the case of the physician analysis, there are several important things to note about this figure. First, the purpose of graphing a non-parametric fitted curve to the data when using the RD design is to get a sense of what the underlying parametric form might be. Although this data does display four points where the slope appears to be zero, suggesting the possibility of a fifth-order polynomial, two of the points are very close together and of relatively small magnitude, suggesting that a third-order polynomial may be an adequate fit. The subsequent analysis includes polynomial estimates up to the quartic form. Second, unlike in the case of the physician data analysis above, data shows what appears to be reasonable density around the cutoff.

Third, also unlike the physician analysis above, the data appears to be relatively balanced on either side of the cutoff without the dramatic skew toward the origin as found in the physician practice size data. As such, this data appears visually to be much more amenable to support stronger claims regarding effect sizes near the cutoff.

#### **7.5.4.2 Assessing the Possibility of Sorting near the Cutoff**

As in the case of the physician data analysis above, the next step in the analysis and presentation of the results of a research question structured using an RD design is to assess whether there appears to be differences in data density immediately on either side of the cutoff that would suggest precise sorting by the units of analysis. Both visual and analytic approaches are available for this test. First, as shown in figure 7.6 on page 133, the data density is plotted against the running variable. (As in the previous figure, the cutoff is indicated with the dotted line.) Also shown is a loess-smoothed curve tracking the density.

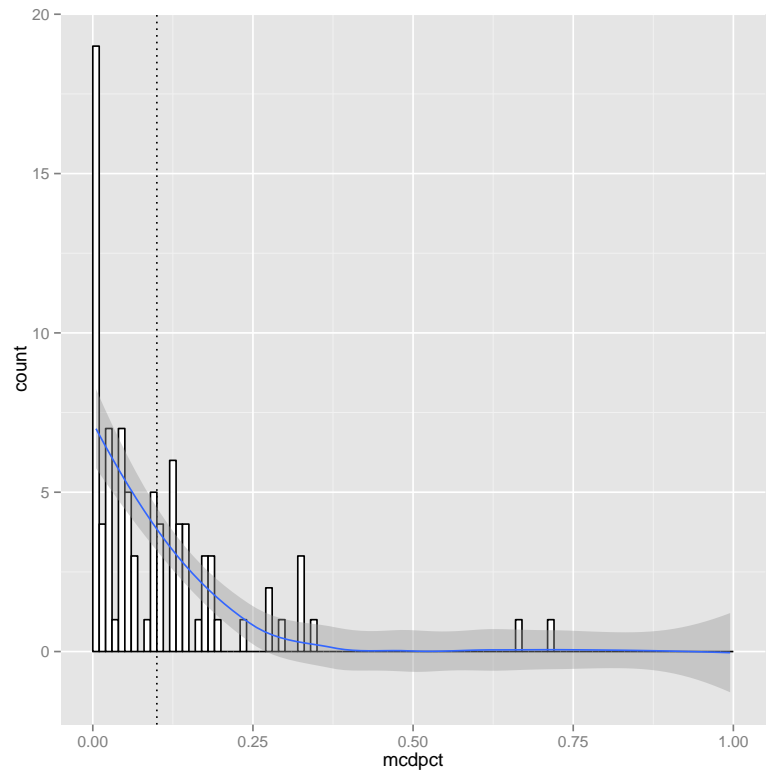


Figure 7.6: Hospital Medicaid % - Histogram, Fitted Curve, and Cutoff

In this case, given that the additional incentive applies when a hospital's Medicaid volume is greater than 10%, if there were sorting near the cutoff, one would expect values immediately above the cutoff to have higher percentages than corresponding values immediately below the cutoff. In this case, there does not appear to be a higher data density just above the cutoff. As in the case of the physician data, a more formal approach would be preferable, but the density comparison calculation suggested by McCrary (2008)<sup>3</sup> fails due to lack of adequate data density around the cutoff.

#### **7.5.4.3 Optimal Bandwidth Calculation**

The next step in the analysis and presentation is to determine if there is an optimal bandwidth for the local linear regression approach. Following the approach suggested by Imbens and Kalyanaraman (2011)<sup>4</sup>, the optimal bandwidth is calculated as 0.1005.

#### **7.5.4.4 Graphing Different Parametric Forms**

As in the analysis of the physician data above, the next step in the analysis and presentation is the graphical display of estimates based on several different bandwidths and parametric forms. Figure I.1 on page 204 shows the graph of linear models with bandwidths 0.05 and 0.10 and polynomial estimates for quadratic, cubic, and quartic functions based on data across the

---

<sup>3</sup>McCrary 2008.

<sup>4</sup>Imbens and Kalyanaraman 2011.

whole of the remaining data space. Figure I.2 on page 205 is a magnified version of the region of figure G.1 outlined by the dotted rectangle.

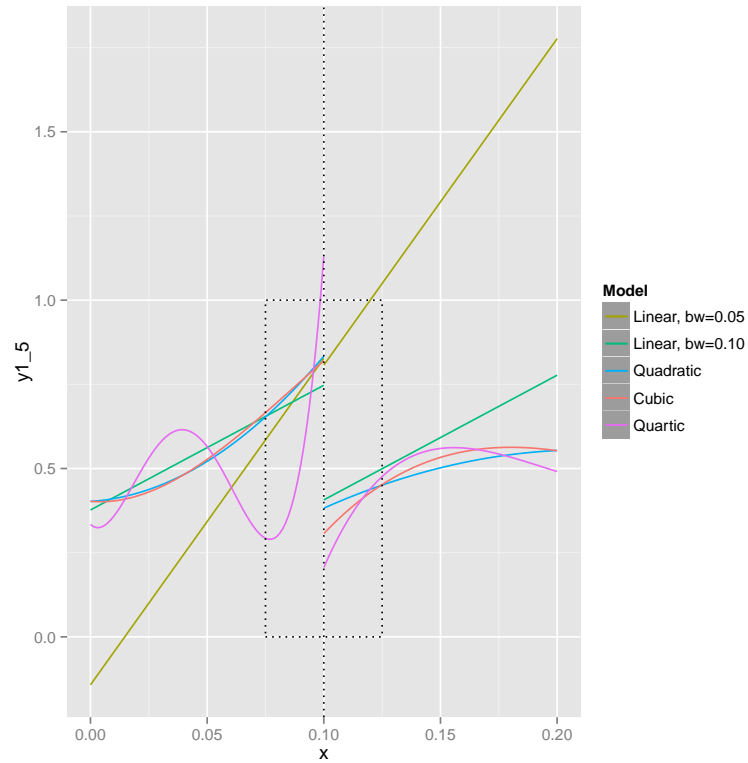


Figure 7.7: MCD RD design Table - Multiple Models

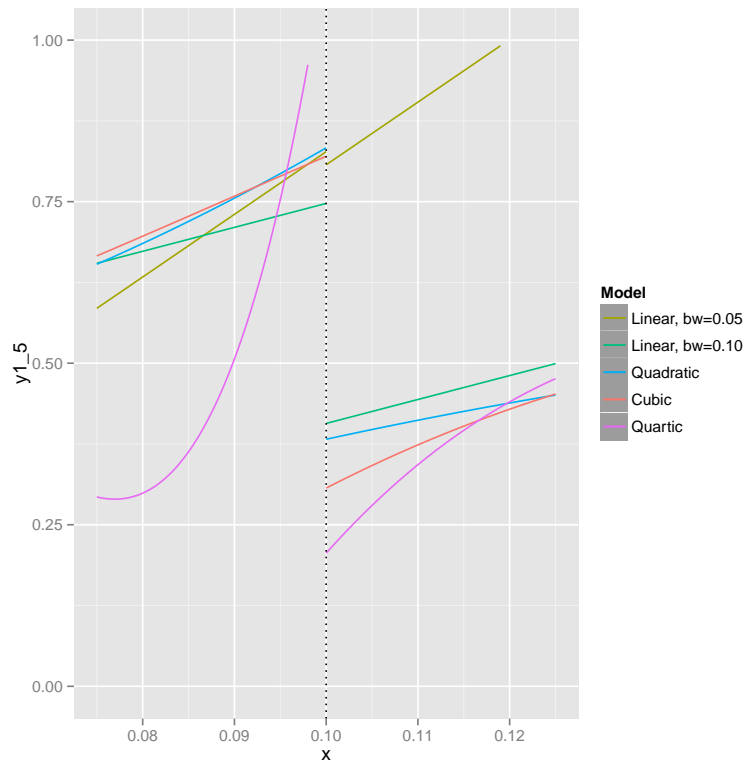


Figure 7.8: MCD RD design Table - Multiple Models, Zoom

As in the case of the analysis of physician data above, the presence of consistent effect estimates and statistical significance across a wide range of specifications supports greater confidence in the estimate.

Figures I.1 and I.1 contain too many different estimates to include confidence intervals for all of them in a way that could be visually discernible so the effect sizes and relative significance levels are summarized in table J.1 on page 207.

	Estimate	Pr(> t )
Local linear, bandwidth=0.05	-0.0203	0.9875
Local linear, bandwidth=0.10	-0.3404	0.5240
Quadratic	-0.4642	0.5994
Cubic	-1.5880	0.5616
Quartic	-5.5750	0.5286

Table 7.2: Medicaid RDD Effect Sizes at Cut-off

### **7.5.5 Discussion**

The Medicaid EHR incentive program for hospitals required for hospitals to have at least 10% Medicaid volume. Hospitals above this threshold may qualify for additional funding for the meaningful use of certified electronic health records, on top of that for which they may also qualify through the corresponding Medicare program. There does not appear to be any theoretical *ex ante* reason to believe that hospitals on either side of the 10% threshold differ in any substantive ways, allowing us to leverage this threshold as the key feature in a regression discontinuity design. Given the conventional wisdom of cost as an impediment to EHR adoption, the natural hypothesis for this program is that it will increase the likelihood of adoption. The results, however, show the opposite, although not at a statistically significant level. There is no clear reason why the availability of additional funds would create a lower adoption rate in this group, although there could be other underlying drivers that are getting picked up by the variables in use.

Appendix I includes additional analysis in which covariates are added to the RD design models discussed above, and appendix J includes additional analysis in which additional bandwidths are tested with all of the parametric forms discussed above.

### **7.5.6 Limitations**

The hospital analysis was particularly limited by the sample size, which would have prevented any but fairly large effects from showing significance.



## 7.6 Conclusion

Certain programmatic aspects of the Health Information Technology Regional Extension Center (REC) program for physicians and the Medicaid EHR incentive program for hospitals lent themselves to analysis via a regression discontinuity (RD) design. The hypothesis in both cases is that eligibility for the relevant program would lead to higher EHR adoption rates. In both cases, the analytical findings were inconclusive, suggesting that the data did not support rejection of the null hypotheses.

## Chapter 8

# Exploring the attitudes and motivations of health care providers regarding electronic health records and federal programs to promote their adoption

### 8.1 Overview

In the preceding chapters, quantitative analyses have been performed to attempt to characterize the impact that several federal programs have had on the adoption of electronic health records (EHRs) by physicians and hospitals in Texas. In this chapter, I will attempt to provide some more qualitative context to this situation by reviewing findings regarding provider attitudes and motivations about EHRs from the previously discussed surveys, as well as information collected from interviews with both EHR-adopting and EHR-non-adopting physicians. I am not attempting, in this chapter, to assert or test any hypotheses, but rather to describe some of the dynamics at the individual and institutional that might be involved with the adoption behaviors. If, in fact, as the data seems to suggest, the federal programs established to promote the adoption of EHRs are not having any measurable effect, an analysis of the attitudes of health care providers towards EHRs may shed some light on the reasons that these programs are not impacting adoption rates.

## 8.2 Data

The observations discussed in this chapter come from three distinct sources. The first source is a set of relatively informal interviews performed by the Public Policy Research Institute at Texas A&M University, under contract with the Texas Health and Human Services Commission, the purpose of which was to gather information about the adoption choices and attitudes of outliers (outliers were defined as the set of physician EHR adopters who had the lowest predicted likelihood of adopting and the set of physician EHR non-adopters who had the highest predicted likelihood of adopting). The second source is the physician surveys used in previous chapters as the basis for quantitative analysis. In addition to capturing details about physician practice characteristics, EHR use, etc, the physician surveys included questions regarding physician attitudes toward EHRs. The third source is the hospital surveys, which also included questions regarding attitudes toward EHRs.

## 8.3 Physician Interviews

Based on an early version of the logit models used in previous chapters, outlier physicians were identified whose predicted adoption likelihood was most contrary to their actual observed behavior. Information was then collected from a small sample (10) of these physicians at each end of the distribution through structured interviews. Of the physicians interviewed who were classified as non-adopters (5), one of them adopted an EHR between the survey and the interview, and another appears to have been in a practice that

partially adopted an EHR. The "adopters" were slightly younger than the "non-adopters", although they had similar self-reported levels of familiarity and comfort with technology, as measured by questions regarding their use of digital technology outside of work and for communication with friends and family.

### 8.3.1 Question Variations

In addition to the questions that were posed in identical form to "adopters" and "non-adopters", several questions were posed in relative forms – retrospective/factual to "adopters" and hypothetical to "non-adopters". For example, "adopters" were asked – "Prior to adopting EHR how difficult or easy did you think that learning to operate EHR would be for you? What do you think informed that perception?" and "How did your expectations compare with your experience? ", while "non-adopters" were asked – "How difficult or easy do you think that learning to operate an EHR would be for you? What do you think informed that perception?". Notably, for this question set and others of similar structure in the interviews, the 'ex ante' perceptions of both "adopters" and "non-adopters", were aligned almost opposite to what might have been expected. Those who had adopted reported prior anticipation of difficulty, while those who had not adopted generally predicted that such a transition would be easy.

### **8.3.2 Perceived Impact of EHR Adoption on Quality of Care**

The interviews included several questions about the physicians' perceptions regarding the anticipated or experienced impact of the EHR on quality of care, including a direct question about the impact on the quality of clinical decision making, and then related questions regarding quality of communication, avoidance of medication errors, and adherence to clinical guidelines. On the questions regarding the impact of the EHR on quality of care, the comments from adopters and non-adopters spanned a similar, very mixed range. In both groups, there were those who believed that EHRs would/did improve the quality of care along multiple axes and those who asserted the opposite.

### **8.3.3 Perceived Impact of EHR Adoption on Patient Interactions**

Another set of questions focused on the impact of EHR adoption on interactions with patients, including the quality of the patient-physician relationship, quality of communication, and patient expectations. The answers to these questions followed a similar pattern, with some respondents on either side of the answers among both the adopters and non-adopters. While there were some negative responses from both sets of physicians, the comments from the non-adopters more frequently included stronger negative perceptions of the impact of the EHR on the patient-physician relationship, generally with emphasis on extra time spent with the EHR and time taken away from the patient encounter.

### **8.3.4 Perceived Financial Characteristics of EHR Adoption**

Another set of questions focused on the finances of EHRs, including questions regarding perceived cost, ROI, financial incentives, and meaningful use. Oddly, the adopters generally seemed to be less confident in their knowledge of EHR costs, even those in solo practice, than non-adopters. Non-adopters seemed to perceive that EHRs were very costly. None of the adopters reported having received financial incentives to adopt EHRs. All of the interviewed non-adopters indicated that financial incentives would encourage them to adopt an EHR, with most indicating that the financial incentives should cover 100% of the costs. The adopters generally indicated an intent to achieve meaningful use, while the non-adopters were mixed in this respect.

## **8.4 Physician Survey**

The physician survey included several questions about physician perceptions of the benefits and characteristics of EHRs. (The following results did not come from the interviews, but rather from the 2012 survey; n=824.) Several of the questions focused on the relationships between EHR use and clinical quality. Figure 8.1 on page 145 shows physician perceptions of the impact of EHRs on several aspects of clinical quality. On each of these measures of perceived EHR impact on clinical quality, a strong majority of physicians who have adopted EHRs expressed that they believed that the EHRs had a positive impact on clinical quality, with the strongest sentiments associated with improvements in communications among the health care team.

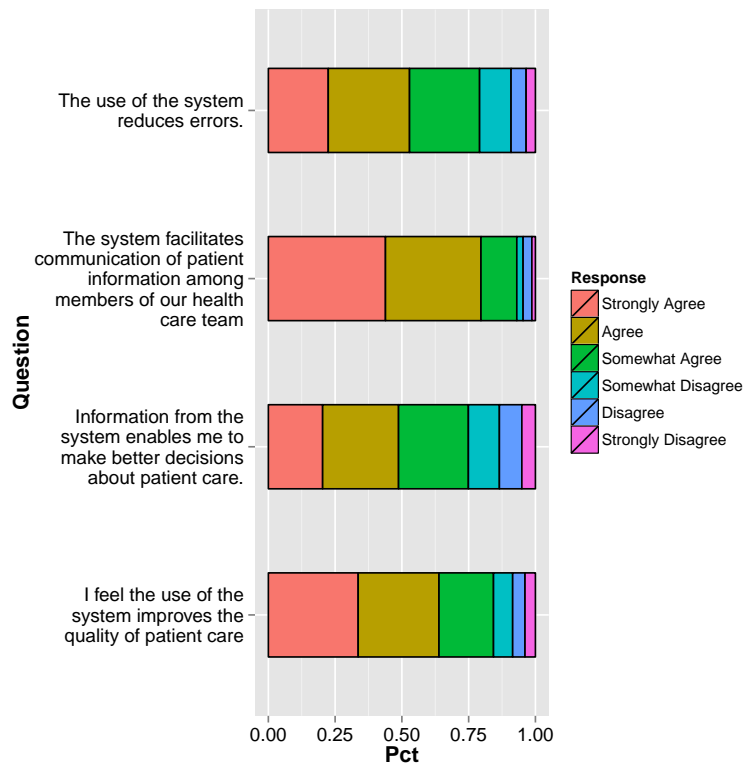


Figure 8.1: Perceptions regarding EHR Impacts on Quality by Physicians who have Adopted EHRs

Similarly, several of the questions in the physician survey focused on the usability of EHRs. Figure 8.2 on page 146 shows physician perceptions of the usability of EHRs. By a significant margin, adopters of EHRs report comfort with using the systems and a preference for EHRs over previous paper records.

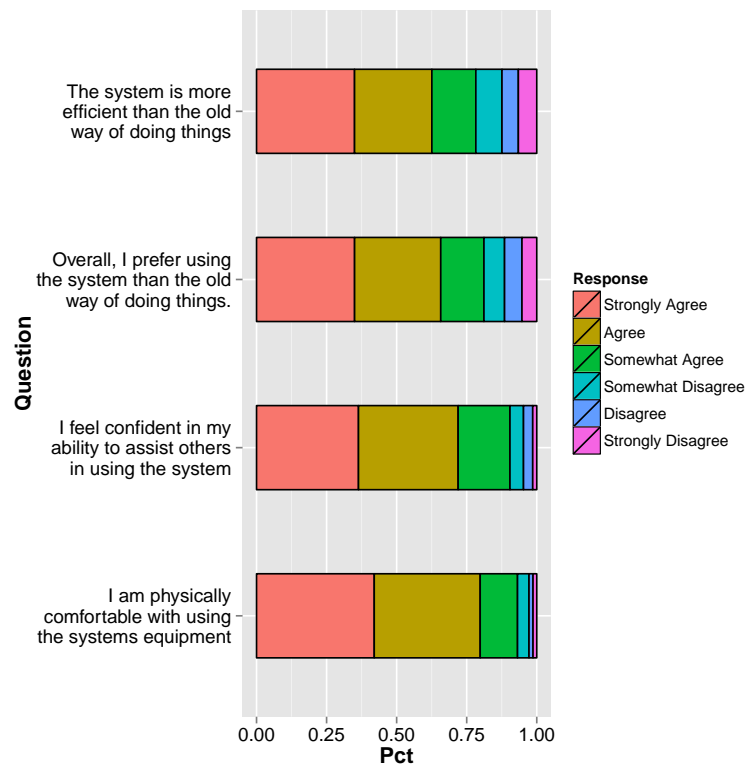


Figure 8.2: Perceptions regarding EHR Usability by Physicians who have Adopted EHRs



In addition to asking EHR adopters how they felt about the use and impacts of EHRs, the survey also asked non-adopters about why they had not yet adopted EHRs. Figure 8.3 on page 147 shows the reasons for not adopting given by physicians who had not yet adopted EHRs. The top reasons given by non-adopters for not adopting EHRs were all related to cost – hardware, software, and personnel costs, and return on investment. Other significant concerns included privacy and security, lack of customization for specific needs, and interoperability.

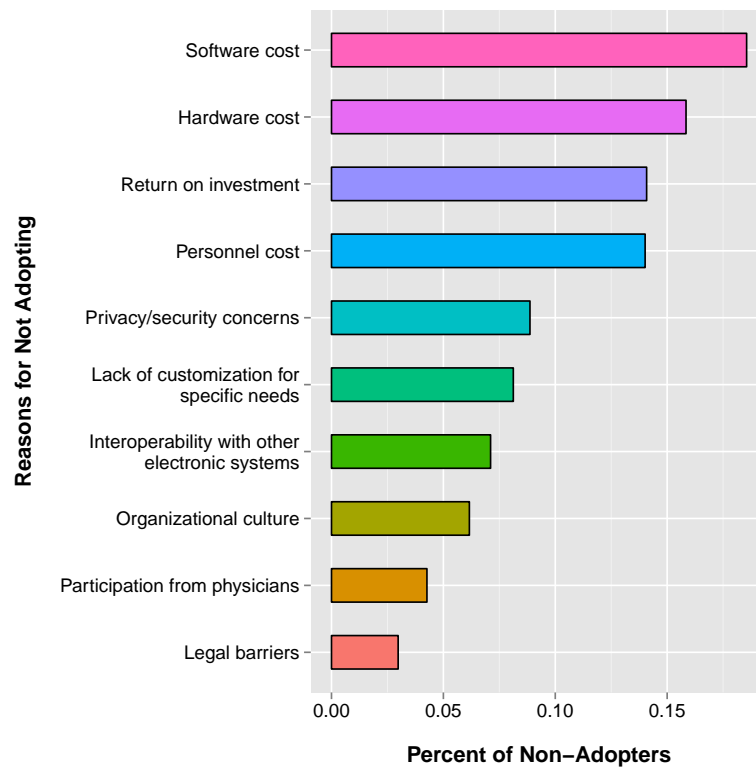


Figure 8.3: Reasons for Not Adopting as Reported by Non-adopters

The physician survey also asked all respondents about whether or not they intended to pursue the financial incentives for the meaningful use of EHRs. Figure 8.4 on page 149 shows physician intentions regarding the meaningful use incentives, broken out by EHR adoption status. Of particular note, although both adopters and non-adopters show relatively similar levels of intent with respect to pursuing the incentives, both levels are well below 50%. This finding could be taken to suggest that the EHR adoption incentives are not particularly compelling if even physicians who have already adopted EHRs are relatively disinclined to pursue the incentives. Alternately, this could be taken as a point of evidence that the choice to adopt an EHR is independent of the EHR incentives.

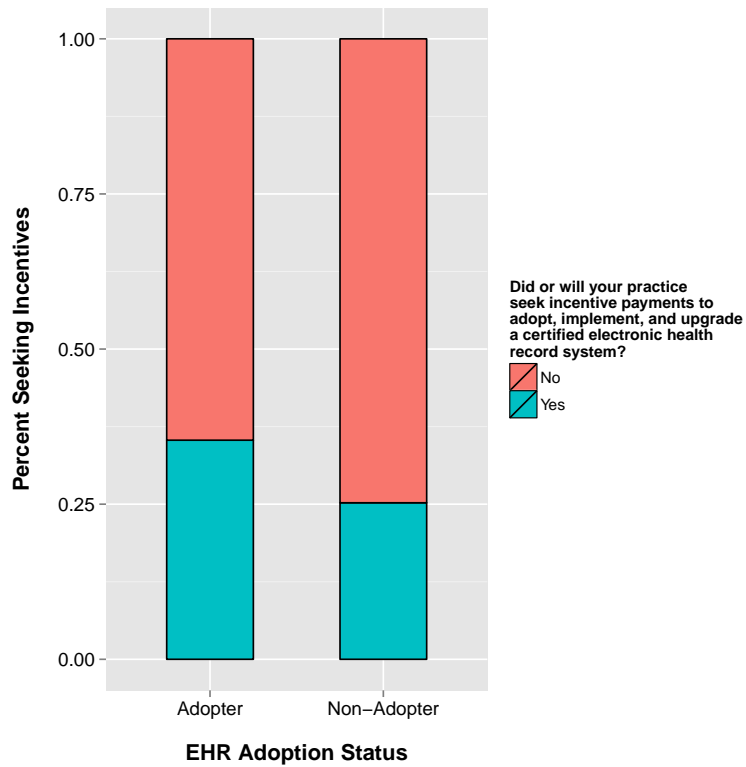


Figure 8.4: Actual or Intended Pursuit of EHR Incentives

## 8.5 Hospital Survey

The hospital survey also included questions for both adopters and non-adopters about perceived benefits and challenges regarding EHR adoption. Figure 8.5 on page 151 shows reasons provided by hospitals that do not have any plans to adopt an EHR.

The hospital sample was numerically smaller than the physician survey by an order of magnitude, although the overall response rate was substantially higher. (The hospital survey included 177 responses of the approximately 600 hospitals in the state, whereas the physician survey included responses from just under 2,000 physicians of the over 50,000 active physicians in the state.) The questions regarding non-adoption in the hospital survey were only given to hospitals that reported having no intentions of adopting. Given the relatively high overall level of adoption among hospitals, and the significant proportion of non-adopters that expressed an intent to adopt, the total number of hospitals reporting no intention of adopting was very low (n=7).

Although the number of responses does not provide strong ground for generalization, the nature of the responses is suggestive as to the possible sentiments of this group. As in the physician survey, hardware and software costs topped the list of concerns. Unlike the physician survey, organizational culture and interoperability were the next highest reasons given for not intending to adopt.

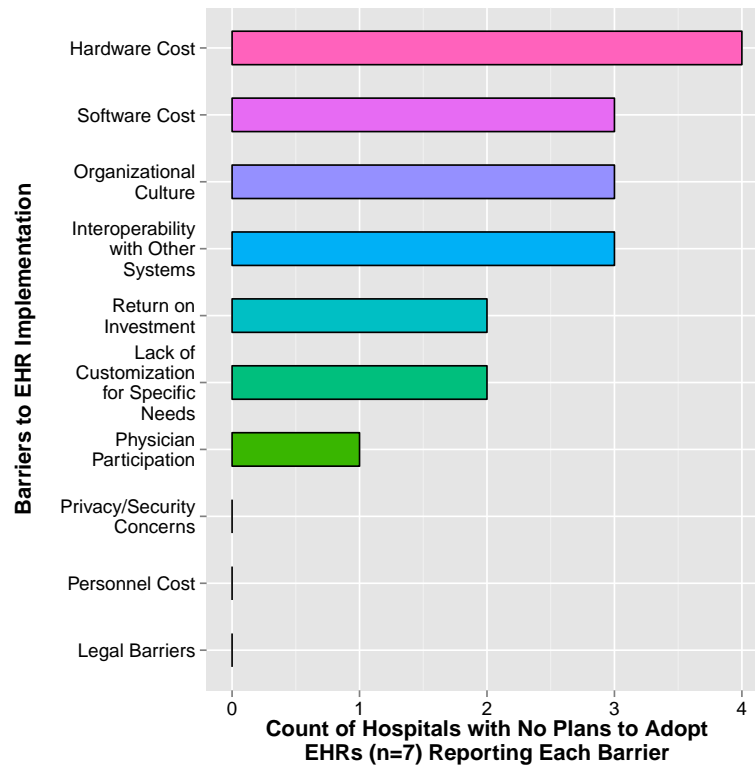


Figure 8.5: Hospital Reasons for not Planning to Adopt EHRs

As with the physician survey, the hospital survey included questions regarding the EHR incentives. Unlike the physician survey, it appears that nearly all hospitals intend to pursue the financial incentives associated with EHR adoption. Figure 8.6 on page 152 shows hospital intentions regarding pursuit of meaningful use incentives, broken out by adoption status. Interestingly, those hospitals who have already adopted EHRs show only marginally higher tendencies to pursue the EHR incentives than those which have not adopted EHRs yet.

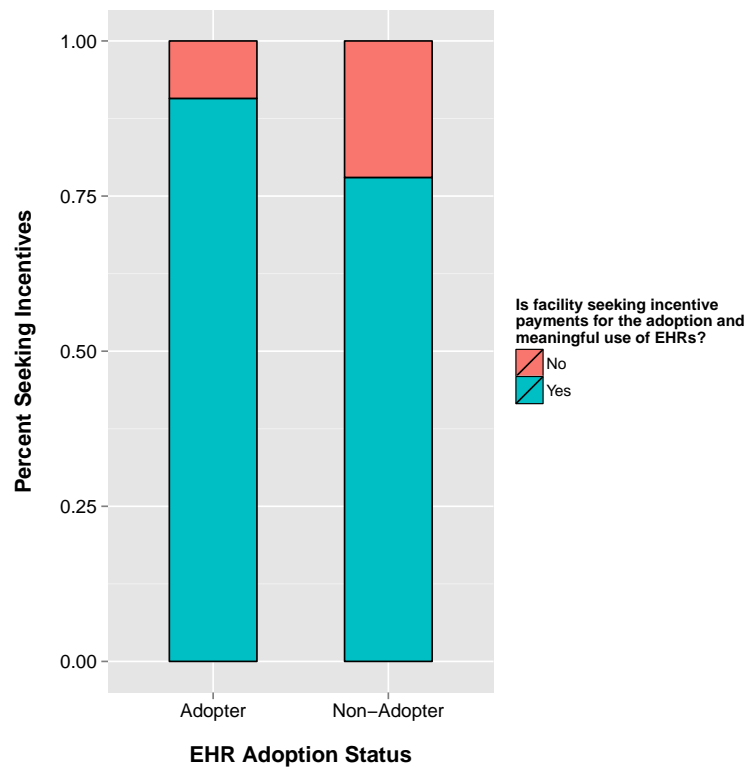


Figure 8.6: Hospital Plans Regarding EHR Incentives

## **8.6 Discussion**

The interview results suggested that both adopters and non-adopters of EHRs have mixed views of the impacts of the EHRs on quality, patient interactions, and practice finances, with few significant differences. The surveys of physicians and hospitals provided a little more detail, but appeared to show adopters and non-adopters to have similar tendencies toward the EHR incentives. Notably, physicians did not appear to be strongly inclined toward the incentives, whether they had already adopted or not, whereas hospitals generally showed a strong inclination toward the incentives, again, whether they had already adopted or not. In both cases, across each surveyed group, the physicians and hospitals appeared to have similar inclinations toward the incentives, whether they had already adopted or not, potentially suggesting that the incentives were not a particularly strong motivator.

## **8.7 Limitations**

While the interviews do provide a little more detail than the surveys, they did not include only limited questions specifically focusing on whether the financial incentives or Regional Extension Centers impacted the physicians' adoption decisions. Likewise, the surveys included some related questions but did not ask specifically about the perceptions of causality or lack thereof between the federal programs and adoption. The generalizability of the information gleaned from the interviews is also limited by their small number.

## 8.8 Conclusion

Taken as a set of findings, the interviews and surveys do not seem to support the contention that the EHR incentives are serving as a motivating factor.



## Chapter 9

### Discussion

Several programs to promote the adoption of EHRs by doctors and hospitals were established by Congress in 2009 through the American Recovery and Reinvestment Act (ARRA). The programs included complementary components through Medicare and Medicaid to provide direct financial incentives for the meaningful use of certified EHRs, and a national program supporting the deployment and activities of local Health IT Regional Extension Centers (HITRECs) to provide direct technical assistance to primary care physicians in small practice groups in the form of professional services to support EHR product selection, workflow re-engineering, and practice optimization.

Prior to the establishment of these programs, hospitals and practitioners used EHRs at rates substantially below the rates of use of IT in other industries. The explicit purpose of these programs was to promote the adoption of EHRs by hospitals and practitioners. As such, the success of these programs would be judged based on whether they, in fact, increased the adoption of EHRs by hospitals and practitioners. In addition, certain features of the EHR promotion programs appear to have been designed in order to address particular disparities in the adoption of EHRs. Therefore, an additional

relevant component of the analysis of these programs would be whether the disparities in the EHR adoption were less after the programs were established than before.

The four analytic approaches pursued in this study attempted to provide different and complementary perspectives on the question of whether the EHR promotion programs have been successful. The first examined whether disparities in EHR adoption rates were reduced after implementation of the programs. The second examined whether the rate of adoption appeared to have changed after implementation of the programs. The third examined whether the programs appeared to have had a marginal effect at certain programmatic thresholds, and the fourth examined whether self-reported information from physician outliers at both ends of the distribution were consistent with anticipated program effects.

## **9.1 Impact of Programs on Patterns of Adoption – Logit Analysis**

Through the first approach – analysis of the patterns of EHR adoption – mixed results were obtained, some of which were consistent with the programs having the desired effect, some of which were not. In particular, the results suggested that after the establishment of the programs, there remained significant disparities in physician EHR adoption rates as a function of practice size. Furthermore, the patterns of adoption did not change in significant ways between the two physician surveys. Importantly, given the structure of these

analyses, no causal claims can be made; the only supportable claims are those of logical consistency (i.e., the results obtained were consistent with a programmatic effect in the anticipated/desired direction.) Similar results were obtained for the hospital models, with the most significant result being the correlation between size and adoption across both surveys. As in the physician models, the patterns in adoption did not change significantly between the surveys.

## **9.2 Impact of Programs on Rates of Adoption – Bass Diffusion Model**

The second approach tested whether the rate of EHR adoption among physicians was greater after the implementation of the programs than before. While the direct result of the analysis was inconclusive, indicating that the rate of adoption was similar before and after the establishment of the programs, drawing broader conclusions about whether this similarity in outcomes was due to a lack of effect from the programs requires making significant assumptions about the counterfactual situation (i.e., what would have happened in the absence of the programs?) While the classic Bass diffusion model does present an interesting and perhaps compelling answer to the counterfactual question, in that it posits a smooth, deterministic adoption path, the reality is often more complex.

As can be seen from real-world adoption curves, while they do often pursue the familiar S-shape of the Bass model, they can be significantly im-

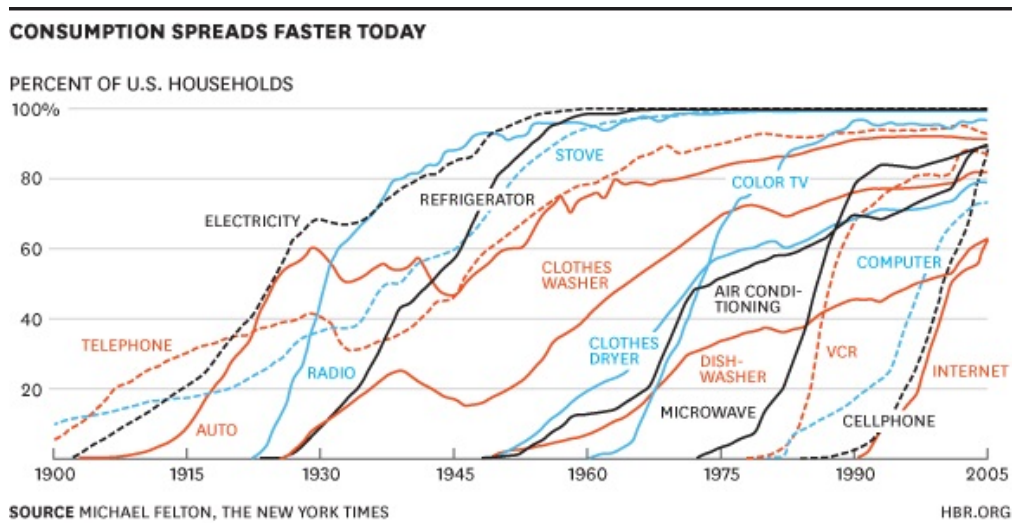


Figure 9.1: Adoption of Household Products in the U.S.

pacted by macro-economic effects, which were not modeled in this analysis. (The graph of different historical adoption curves is included again as figure 9.1 on page 158 for reference.) For example, observing the adoption curve of automobiles across the time period of the Great Depression, it became jagged and declined. Although the clothes dryer enjoyed a strong positive adoption curve during the Great Depression, its adoption curve dips during World War II, when a strong demand for raw materials probably created a supply-side constraint. Considering the possibility of macro-economic effects on EHR adoption, particularly in light of the reality that these programs were established during a major macro-economic recession, as part of a piece of legislation, the stated purpose of which was to stimulate the economy, it may be the case that the programs functioned exactly as envisioned, to stimulate both supply

and demand. Observing the downward trends in a number of the historical adoption curves during macro-economic downturns, against the observation of the monotonically positive EHR adoption curve, provides some support to the claim that the EHR promotion programs had a positive effect.

### **9.3 Impact of Programs on Adoption Levels – RD Design**

The third approach attempted directly to measure the impact of the programs on the adoption levels by using regression discontinuity (RD) designs. The models based on RD designs did not identify a significant impact of the REC program on adoption by primary care physicians, nor did it find a significant impact of the Medicaid EHR incentive program on hospital EHR adoption. This approach did not attempt to capture all possible mechanisms by which the programs as a set affected EHR adoption, nor are these findings generalizable substantially beyond the cutoff points analyzed in each case. One element of each of the physician and hospital programs was tested using this approach. The models based on RD designs for physicians specifically tested whether there was an effect of the REC program at the 10 physician practice size threshold and the analysis of models based on RD designs for hospitals specifically tested whether there was an effect at the 10% Medicaid caseload threshold. In neither case was the effect significant at the cut-off (the relevant measure for models based on RD designs.)

RD designs are generally considered to be strong research designs due

to their similarity to randomized controlled experiments. This similarity and, hence, the strength of the approach for supporting generalizable causal inference only holds if the cutoff is observed strictly. While it is the case that both of these threshold measures are strictly enforced programmatically, it could be the case that providers near the cutoff might be able to game the program constraint in order to get the benefit of the programs. For example, a primary care physician group above but near the 10 member cut-off for technical support under the HITREC program might be able to claim that it, in fact, had only 10 members by attributing members to other groups, breaking the group into multiple groups but maintaining common administrative and business services, or other corporate structure gyrations. Likewise, a hospital with just under 10% Medicaid case volume might take on additional Medicaid caseload in order to get over the 10% threshold. In fact, since eligibility for the Medicaid incentive payments is binary at the threshold, then there would be some amount under the threshold that it would be economically rational to do exactly that, even if doing so crowds out better payers at the margin.

## **9.4 Limitations**

In addition to the limitations to each of the individual analyses discussed above, there are several additional limitations to these analyses that are not specific to one or another of them, particularly with respect to our ability to make generalizable causal inferences from them. General issues with these analyses include low survey response rates, the possibility of lagged pro-

gram effects, the dependence on theory to establish a counterfactual, and the possibility of instrumentation bias.

The response rate in the physician survey was low, raising the possibility that it is not representative of the universe of all physicians in Texas, even given strong similarity between the sample and the set of all physicians in the state with respect to certain observable characteristics. There is always the possibility that the sample differs from the set of all Texas physicians based on some unobservable but relevant characteristics.

In addition, the effect from the EHR promotion programs may be lagged. Although the programs were authorized in early 2009, the first payments were not made until 2011. The REC program began enrolling providers in late 2009, but did not gain momentum until late 2010. The data analyzed in this report was collected between 2010 and 2012, so the effects from the EHR promotion programs may have not yet had a chance to trickle through into physician and hospital behavior. On the other hand, it seems as if the knowledge that the incentive payments could be available would have been adequate to induce behavior change since physicians and hospitals tend to have easy access to capital.

In the case of the Bass diffusion analysis, the counterfactual against which the observed adoption levels were measured was heavily informed by theory. Thus, any finding that the observed values differed in a statistically significant way from the projected values is highly dependent on the validity of the underlying theory. Since actual technology adoption curves often deviate

from the theoretical projection, these findings rest heavily on the strength of the theory, which often does not exactly predict real-world empirical results.

Finally, questions about the adoption of EHRs are notoriously fickle instruments. One source of the fragility of EHR questions on surveys is that it often might not be physicians themselves providing the answers. Nurses, medical assistants, or administrative staff may answer for physicians, and may not be as familiar with the terminology being used. When attempting to ascertain physician attitudes and/or behaviors through the use of surveys, physicians themselves are often not available. Surveys are sent by mail to physician offices, sent by email to physician email addresses, or fielded by phone, with calls going to physician office phone numbers. While it is certainly possible in principle to insist that only the physician responds to the survey, in practice doing so would likely drive an even lower response rate. In addition, even if physicians themselves answer the questions, their awareness of the terminology may not be any greater than that of their staff.

## **9.5 Review of Key Findings**

- Provider size is highly negatively correlated with physician and hospital EHR adoption, with no significant change in this pattern after the implementation of the EHR promotion programs.
- The physician EHR adoption rate is not substantially different than what diffusion of innovation theory suggests would have been the case in the



absence of the EHR promotion programs.

- The Medicaid EHR incentive program for hospitals did not appear to have a statistically significant effect on hospital EHR adoption levels near the 10% Medicaid caseload threshold.
- The HITREC program did not appear to have a statistically significant effect on primary care physician EHR adoption levels near the 10 physician practice size threshold.

Taken as a set, the findings from the several lines of analysis pursued in this study are sufficiently weak as to prevent making any defensible causal claims.

## **9.6 Policy Implications**

Taken as a set, the findings from this research are not consistent with the claim that the EHR promotion programs had any significant effect on physician or hospital EHR adoption as of 2012. Although there are certainly numerous limitations in this study that might give reason to question the validity and/or generalizability of these findings, it is still useful to consider whether the public investment of funds is worth it if we take these results as given. From the beginning of the EHR incentive programs in January 2011 through August 2014, just over \$1.8 billion in Medicaid and Medicaid EHR incentive payments had been made to practitioners and hospitals in Texas alone for the meaningful use of certified EHR technology, with a total national spend of nearly \$25 billion over the same period. In addition, over \$500 million

was spent nationally on the HITREC program, with about \$35 million of those funds being allocated to the four Texas RECs.

In the two years since the data analyzed in this dissertation were collected, EHR adoption rates have continued to rise among both physicians and hospitals. While the promoters of the original legislation and public administrators charged with implementing it have touted these increased EHR adoption rates as signs of the programs' successes, their claims have been little more than observations of the correlation. Given the nationwide nature of the programs, the lack of an empirical counterfactual, and the absence of strong research into the question of the programs' effectiveness, it appears very unlikely that anybody would be justified in making strong causal claims at this point.

Even if EHR adoption levels by physicians and hospitals would have been at a similar level absent the EHR promotion programs, there may be some aspects of these programs that have contributed to broader health policy goals. For example, that all of the financial incentives paid through these programs have required providers to demonstrate the meaningful use of the technology may help to nudge providers toward more efficient and effective clinical practice. And the requirement that providers can only receive the financial incentives for using certified EHRs provides a potential mechanism for ensuring that the different systems in use are able to communicate with each other, often referred to as 'interoperability'. Although interoperability might have been a key outcome of the EHR certification process, the federal agen-

cies tasked with implementing these programs have received criticism for not achieving greater interoperability through the first two phases of meaningful use and certification. However, recent developments suggest that the responsible federal agencies may attempt to rectify this shortcoming through the third phase of meaningful use and certification requirements. It is probably too early to determine if the meaningful use requirements have led to detectable changes in clinical practice, and this does leave us with the question of whether the significant federal investment was worth it if interoperability remains elusive.

Another key implication of these findings and any like them is on the prospects of additional federal HIT funding. Soon after ARRA, which included EHR adoption incentives for hospitals and certain practitioners, was passed, it was quickly noted by other health care communities that some types of health care providers were left out of the incentives, notably behavioral health and long-term care. Negative findings regarding the impacts of the existing EHR promotion programs would probably make Congress less likely to authorize additional funding for HIT in other contexts.

## **9.7 Market Analysis**

In consideration of the possibility that the EHR promotion programs included in ARRA had no marginal effect, one interesting angle is the review of stock prices for publicly traded EHR companies. Considering the efficient market hypothesis, all relevant information that might affect the future sales of the underlying companies should affect the share price immediately. Within

the set of inpatient and ambulatory EHR companies, few have been publicly traded as EHR-only companies over the period spanning several years prior to ARRA through the present. Among the set of inpatient and ambulatory EHR companies, most have been subject to corporate activities or have corporate structures that inhibit a clear analysis of their finances. For example, one of the more popular ambulatory EHRs is made by GE, but the impact of the GE EHR on the GE share price must be assumed to be negligible – likewise the hospital EHR product until recently owned by Siemens. In addition, some EHR companies have been subject or object of significant acquisitions, sometimes of companies that are not directly in the EHR space, muddling the information that can be gleaned from publicly available stock prices. Finally, some EHR companies are currently or have always been privately held, either by founders or private equity groups.

Reviewing the set of EHR companies in both the inpatient and ambulatory spaces that are currently publicly traded, have been publicly traded for several years prior to ARRA, and are not within conglomerates that obfuscate the role of the EHR offerings on the share price, two companies stick out – Cerner in the inpatient space and AthenaHealth in the ambulatory space. Figure 9.2 on page 167 shows the percentage changes in these two companies' stock prices from more than a year prior to ARRA through the present. The two stock prices appear to be fairly closely correlated, and while both have experienced over 200% growth over the last 6 years, neither experienced a dramatic uptick in early 2009 at the time of the passage of ARRA. In fact, the

growth in the values of both stocks has been pretty steady over the last five years. It seems that, if in fact the collective wisdom of the market perceived that the passage of ARRA would lead to a substantial marginal increase in EHR adoption rates, then these stocks should have seen a substantial upward around the time of passage. Contrary to this possibility, it appears that the values of these stocks actually just increased as these companies organically increased their sales.



Figure 9.2: Stock Prices for Two Leading EHRs

## 9.8 Extensions

There are several key ways that this research might be extended or improved. In particular, having a better response rate on the physician survey, a larger sample on the hospital survey, more explicit questions on the surveys and interviews regarding the perceived motivating influence of the EHR promotion programs, an examination of clinical impacts of EHR adoption, more explicit questions and analysis regarding the role of new payment models and

hospital alignment, and better qualitative understanding of the strategies used by the RECs could lead to more useful and generalizable results.

As previously noted, the low response rate on the physician survey draws into question the generalizability of the results. A higher response rate would mitigate this critique.

Similarly, the relatively small number of hospitals in the core survey, while representing a larger response rate, weakened what could have been one of the methodologically stronger analyses in this study, namely the hospital RDD analysis of the Medicaid EHR incentive program. Particularly given the strength of the RDD approach for generating generalizable causal inference, but recognizing the need to have an adequate number of observations very close the cut-off, a larger, probably national sample may be necessary.

Working through the analysis, it became clear that one of the core issues was the question of whether the EHR promotion programs motivated providers to act differently than they would have otherwise. Although the interview questions provided some tentative answers to this question, it was largely implied. More explicit survey questions and deeper interviews on the effect of the EHR promotion programs on the adoption decision would be useful.

Since, based on this analysis, it remains an open question whether the EHR promotion programs had an effect on adoption levels, greater analysis of the other potential effects of the programs could be relevant. In partic-

ular, if one of the possibilities is that the meaningful use requirements resulted in better clinical practices or outcomes, then these measures could be analyzed. Likewise, once interoperability is more strongly promoted through the provider meaningful use requirements and EHR certification requirements, then interoperability and its presumed other clinical effects could be measured and analyzed.

The role of new payment models on both physician and hospital EHR adoption could be analyzed through more specific survey and interview questions. Likewise, the role of physician-hospital alignment on physician EHR adoption could be analyzed. Both of these dynamics may contribute to higher EHR adoption rates by providers.

For the REC program, a somewhat different line of additional analysis may be warranted. The RECs received payment for each provider registered with them that achieved different milestones, including registration, EHR selection and adoption, and meaningful use, they may have had an incentive to pursue providers who were on the cusp of adoption already (i.e., the "low hanging fruit"). As such, if the RECs primarily pursued those providers who were on the cusp of adopting already, then they may have been able to take credit for helping providers achieve meaningful use who would have done so without any assistance. Greater qualitative analysis of the REC strategies for targeting physicians could help to illuminate whether the targeting and recruitment strategies of the RECs may have allowed them to take credit for EHR adoptions that would have happened anyway

## Chapter 10

### Conclusion

In 2009, several programs were established by Congress through ARRA with the express purpose of increasing the levels of EHR adoption by hospitals and certain health care professionals. These programs included financial incentives for the meaningful use of certified electronic health record (EHR) technology by hospitals and certain health care professionals and direct technical assistance in the form of professional services for primary care physicians practicing in small groups. The EHR financial incentives were administered by and disbursed through the federal Medicare program and state Medicaid programs. The technical assistance services were administered by a network of newly established Health Information Technology Regional Extension Centers (HITRECs).

This study attempted to assess the preliminary impact of these federal programs established to promote the adoption of EHRs by analyzing patterns of adoption, rates of adoption, and certain direct impacts of the programs.

Patterns of adoption were analyzed using logistic regression analysis of the characteristics correlated with certain perceived undesirable disparities in the adoption patterns. The logistic regression analysis did not show a



statistically significant change in the patterns of adoption between the two physician surveys, nor did it identify a statistically significant change in the patterns of adoption between the two hospital surveys. These results suggest that, to the degree that certain aspects of the EHR promotion programs were designed to mitigate specific disparities in adoption levels, they did not do so. In addition to this analysis not being structurally capable of supporting causal claims, there are possible sources of bias and threats to validity that could draw the reliability of observed correlations into question.

Rates of EHR adoption were analyzed by fitting annual measures of EHR adoption to technology adoption curves using the Bass formalism of the Rogers diffusion of innovation theory. Predictions from a time series including data before the implementation of the EHR promotion programs were compared to measures of actual EHR adoption after the establishment of the programs to estimate whether the actual measurements varied from the theoretically-supported counter-factual based only on the time series prior to the establishment of the programs. Although two of the actual measures after the establishment of the programs was just outside the projected confidence intervals, the others were not, and the overall findings were inconsistent. Therefore, I can assert that, within the assumptions establishing the counter-factual, the programs did not have a consistent, significant effect. However, this finding is not very strong since there are major assumptions involved in the counter-factual and there are other characteristics of the survey that do not support strong generalizability, such as potential sampling bias.

The direct effects of the EHR promotion programs were analyzed using regression discontinuity (RD) designs through which certain features of the EHR promotion programs could be exploited in order to make direct estimates of the effects of two of the programs, at points near certain programmatic thresholds. In particular, the Medicaid EHR incentive program was tested around the 10% Medicaid volume program eligibility threshold and the HITREC program was tested around the 10 physician practice size program eligibility threshold. RD is generally considered to be a very strong research design since, under assumption of strong adherence to the cutoff, it approximates the circumstances of a randomized controlled trial with all of its concomitant strengths with respect to validity and generalizability. In neither the case of the Medicaid EHR incentive program, nor the HITREC program did the results of the models using RD designs show any significant effect around the cutoff.

Overall, some of the structures of the analyses pursued herein were capable of producing relatively strong, generalizable results, at least within the assumptions relating to the samples and instrumentation. However, given the analytic results, no real claims can be made about the effects of the programs. On balance, and taking an appropriately conservative approach to the interpretation of the findings, it can only be said that they do not support the claim that the programs had an effect.

## Appendices

## Appendix A

### Physician Logit Analysis Including Hospital Referral Regions

The models below use the same variables that were included in the physician EHR adoption logit models with discrete variables, with the addition of fixed effects variables for hospital referral regions (HRRs). This layer of variables is tested to determine if there might be some regional variations. None of the coefficient estimates for the HRR variables were statistically significant.

The following tables show the physician logit models, using factor variables, with the inclusion of HRR fixed effects. The most notable difference between these models and those without the HRR fixed effects is that one HRR (anchored in Odessa) appears to have a positive effect on the physician EHR adoption rate. However, given the number of HRRs (22) and the level of statistical significance indicated ( $p < 0.1$ ), it seems more likely that this is merely an artifact of the number of different HRRs tested.

	<i>Dependent variable:</i>		
	EHR		
	(1)	(2)	(3)
MEDICAID1	0.25*	0.25	0.23
practcnt_lvlsSmall (2-5)	0.73***	0.73***	0.72***
practcnt_lvlsMedium (6-10)	1.14***	1.14***	1.18***
practcnt_lvlsLarge (11-100)	1.62***	1.61***	1.66***
practcnt_lvlsVery Large (101-1000)	2.68**	2.57**	2.58**
zippovpct_lvlsBetween 10 and 20		0.07	0.13
zippovpct_lvlsBetween 20 and 30		-0.12	-0.02
zippovpct_lvlsBetween 30 and 40		-0.02	0.05
zippovpct_lvlsBetween 40 and 50		0.96**	0.95**
zippovpct_lvlsBetween 50 and 60		-1.26	-1.12
zippovpct_lvlsBetween 60 and 70		-0.58	-0.47
MetroStatusMicropolitan			-1.39***
MetroStatusSmall metro			-0.46
MetroStatusMedium metro			0.52
MetroStatusLarge fringe metro			-0.69
MetroStatusLarge central metro			-0.85**
survey_yr	0.49	0.49	-0.33
hrrnum219	1.17	1.24	1.77
hrrnum382	0.34	0.39	0.40
hrrnum383			
Observations	1,871	1,862	1,862
Log Likelihood	-1,179.82	-1,172.56	-1,160.43
Akaike Inf. Crit.	2,475.65	2,485.12	2,480.86

*Note:*

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Table A.1: Physicians, Both Years, Factor Models with HRRs, Table 1

<i>Dependent variable:</i>			
EHR			
	(1)	(2)	(3)
hrrnum385			0.36
hrrnum386			-0.95
hrrnum388			-0.38
hrrnum390			-0.19
hrrnum391	0.19	0.20	0.58
hrrnum393	0.81	0.82	-0.13
hrrnum394	-0.27	-0.56	-0.09
hrrnum396	0.86	0.91	0.09
hrrnum397	0.30	0.32	0.72
hrrnum399	-0.14	-0.12	-0.37
hrrnum400	-0.12	-0.02	-0.42
hrrnum402	0.43	0.46	-0.45
hrrnum406	1.66*	1.74*	1.76*
hrrnum411	1.26	1.28	1.09
hrrnum412	0.34	0.35	0.74
hrrnum413	14.78	14.75	14.05
hrrnum416	-0.24	-0.22	-0.01
hrrnum417	-0.26	-0.15	-0.16
hrrnum418	0.76	0.78	0.68
hrrnum420			
Observations	1,871	1,862	1,862
Log Likelihood	-1,179.82	-1,172.56	-1,160.43
Akaike Inf. Crit.	2,475.65	2,485.12	2,480.86

*Note:* \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Table A.2: Physicians, Both Years, Factor Models with HRRs, Table 2

	<i>Dependent variable:</i>		
	EHR		
	(1)	(2)	(3)
MEDICAID1:survey_yr			0.32
practcnt_lvlsSmall (2-5):survey_yr			-0.04
practcnt_lvlsMedium (6-10):survey_yr			0.15
practcnt_lvlsLarge (11-100):survey_yr			0.17
practcnt_lvlsVery Large (101-1000):survey_yr	-0.83	-1.00	-1.04
zippovpct_lvlsBetween 10 and 20:survey_yr		-0.03	0.04
zippovpct_lvlsBetween 20 and 30:survey_yr		-0.004	0.09
zippovpct_lvlsBetween 30 and 40:survey_yr		0.30	0.43
zippovpct_lvlsBetween 40 and 50:survey_yr		-1.32*	-1.21*
zippovpct_lvlsBetween 50 and 60:survey_yr		0.27	0.31
zippovpct_lvlsBetween 60 and 70:survey_yr		0.25	0.29
MetroStatusMicropolitan:survey_yr			1.18
MetroStatusSmall metro:survey_yr			0.73
MetroStatusMedium metro:survey_yr			-0.14
MetroStatusLarge fringe metro:survey_yr			1.47**
MetroStatusLarge central metro:survey_yr			1.13*
survey_yr:hrrnum219	-16.80	-16.76	-16.83
survey_yr:hrrnum382	0.23	0.26	0.30
survey_yr:hrrnum383	0.64	0.69	0.75
survey_yr:hrrnum385			
Observations	1,871	1,862	1,862
Log Likelihood	-1,179.82	-1,172.56	-1,160.43
Akaike Inf. Crit.	2,475.65	2,485.12	2,480.86

*Note:*

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Table A.3: Physicians, Both Years, Factor Models with HRRs, Table 3

	<i>Dependent variable:</i>		
	EHR		
	(1)	(2)	(3)
survey_yr:hrrnum386			0.96
survey_yr:hrrnum388			1.49
survey_yr:hrrnum390			-0.43
survey_yr:hrrnum391			-0.59
survey_yr:hrrnum393			-0.65
survey_yr:hrrnum394			-0.30
survey_yr:hrrnum396			-1.66
survey_yr:hrrnum397			-0.98
survey_yr:hrrnum399			1.10
survey_yr:hrrnum400	-0.36	-0.40	0.02
survey_yr:hrrnum402	-0.56	-0.63	0.15
survey_yr:hrrnum406	-0.91	-0.97	-0.93
survey_yr:hrrnum411	-1.56	-1.54	-1.38
survey_yr:hrrnum412	-0.17	-0.19	-0.58
survey_yr:hrrnum413	0.18	0.31	0.82
survey_yr:hrrnum416	-0.29	-0.29	-0.36
survey_yr:hrrnum417	-0.41	-0.43	-0.39
survey_yr:hrrnum418	0.01	-0.004	0.25
survey_yr:hrrnum420	14.30	14.32	14.36
Constant			
Observations	1,871	1,862	1,862
Log Likelihood	-1,179.82	-1,172.56	-1,160.43
Akaike Inf. Crit.	2,475.65	2,485.12	2,480.86

*Note:* \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Table A.4: Physicians, Both Years, Factor Models with HRRs, Table 4



## Appendix B

### Physician EHR Adoption Estimated Using Linear Probability Models

As a further test of the robustness of the coefficient estimates, the same models are estimated below as linear probability models. Under these specifications, the same coefficient estimates show similar levels of significance.

	<i>Dependent variable:</i>		
	EHR		
	(1)	(2)	(3)
Medicaid Provider	0.07**	0.07**	0.05
Practice Size - Small	0.17***	0.16***	0.16***
Practice Size - Medium	0.25***	0.25***	0.25***
Practice Size - Large	0.38***	0.37***	0.37***
Practice Size - Very Large	0.56***	0.53***	0.52***
Poverty Pct (10-20)		0.02	0.03
Poverty Pct (20-30)		-0.01	0.01
Poverty Pct (30-40)		0.04	0.03
Poverty Pct (40-50)		0.09	0.11
Poverty Pct (50-60)		-0.30	-0.22
Poverty Pct (60-70)		-0.17	-0.14
Metro-Micro			-0.32***
Metro-Small			-0.12
Metro-Medium			-0.07
Metro-Large Fringe			-0.12
Metro-Large Central			
Observations	1,871	1,862	1,862
Log Likelihood	-1,274.10	-1,266.28	-1,250.88
Akaike Inf. Crit.	2,572.19	2,580.56	2,569.76

*Note:*

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Table B.1: Physicians, Both Years, Linear Probability Factor Models, Explanatory Variables

	<i>Dependent variable:</i>		
	EHR		
	(1)	(2)	(3)
Survey Year			-0.13
Medicaid Provider*Year			0.06
Practice Size - Small*Year			0.01
Practice Size - Medium*Year			0.03
Practice Size - Large*Year	-0.004	0.02	0.02
Practice Size - Very Large*Year	-0.15	-0.18	-0.20
Poverty Pct (10-20)*Year		0.001	0.01
Poverty Pct (20-30)*Year		-0.02	-0.001
Poverty Pct (30-40)*Year		-0.01	0.05
Poverty Pct (40-50)*Year		-0.21	-0.20
Poverty Pct (50-60)*Year		0.09	0.04
Poverty Pct (60-70)*Year		0.07	0.09
Metro-Micro*Year			0.23
Metro-Small*Year			0.24*
Metro-Medium*Year			0.05
Metro-Large Fringe*Year			0.24*
Metro-Large Central*Year			
Observations	1,871	1,862	1,862
Log Likelihood	-1,274.10	-1,266.28	-1,250.88
Akaike Inf. Crit.	2,572.19	2,580.56	2,569.76

*Note:*

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Table B.2: Physicians, Both Years, Linear Probability Factor Models (cont.), Year Effects

## Appendix C

### Hospital Logit Analysis Including Hospital Referral Regions

The models below use the same variables that were included in the hospital EHR adoption logit models with discrete variables, with the addition of fixed effects variables for hospital referral regions (HRRs). This layer of variables is tested to determine if there might be some regional variations. None of the coefficient estimates for the HRR variables were statistically significant.

	<i>Dependent variable:</i>		
	ehr		
	(1)	(2)	(3)
sizeMedium (101-400)	0.82**	0.92**	0.84**
sizeLarge (400+)	1.69***	1.92***	2.27***
hrrnum382	0.53	-0.32	-0.38
hrrnum383	-1.12	-1.53	-1.70
hrrnum385	0.11	-0.45	0.48
hrrnum386	-0.14	-1.35	-1.39
hrrnum388	18.13	17.86	17.74
hrrnum390	-16.74	-18.01	-17.96
hrrnum391	-0.18	-0.92	-0.99
hrrnum393			
Observations	379	379	377
Log Likelihood	-205.16	-195.84	-186.11
Akaike Inf. Crit.	506.33	507.68	508.21

*Note:* \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Table C.1: Hospitals, Both Years, Factor Models with HRR, Table 1

	<i>Dependent variable:</i>		
	ehr		
	(1)	(2)	(3)
hrrnum394	-0.06	-0.52	-0.62
hrrnum396	-0.44	-1.82	-1.85
hrrnum397	0.26	-0.39	-0.52
hrrnum399	-0.78	-0.82	-1.10
hrrnum400	0.36	-0.78	-0.91
hrrnum402	0.99	-0.34	-0.31
hrrnum406	-0.51	-0.90	-0.97
hrrnum411	1.52	0.49	0.43
hrrnum412	0.04	-0.73	-0.81
hrrnum413	0.96	-0.19	-0.38
hrrnum416	1.05	0.44	0.34
hrrnum417	-0.94	-1.33	-1.33
hrrnum418	-0.36	-1.06	-1.08
hrrnum420	-0.43	-1.16	-1.33
urbrurcode2		0.25	0.09
urbrurcode3		0.85	0.72
urbrurcode4		-0.52	-0.56
urbrurcode5		0.75	0.76
urbrurcode6		0.61	0.56
zippovpct_lvlsBetween 10 and 20			
Observations	379	379	377
Log Likelihood	-205.16	-195.84	-186.11
Akaike Inf. Crit.	506.33	507.68	508.21
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01		

Table C.2: Hospitals, Both Years, Factor Models with HRR, Table 1

	<i>Dependent variable:</i>		
	ehr		
	(1)	(2)	(3)
zippovpct_lvlsBetween 20 and 30			-0.09
zippovpct_lvlsBetween 30 and 40			0.11
zippovpct_lvlsBetween 40 and 50			-0.37
zippovpct_lvlsMore than 50			-19.09
survey_yr	0.41	1.44	-0.47
sizeMedium (101-400):survey_yr	0.003	0.55	0.58
sizeLarge (400+):survey_yr	0.92	0.76	0.24
hrrnum382:survey_yr	1.60	2.33	2.94
hrrnum383:survey_yr	1.58	1.10	1.74
hrrnum385:survey_yr	0.22	-1.06	-0.49
hrrnum386:survey_yr	-1.45	-16.48	-15.98
hrrnum388:survey_yr			
hrrnum390:survey_yr	16.92	1.02	1.08
hrrnum391:survey_yr	0.18	-0.71	0.42
hrrnum393:survey_yr			
hrrnum394:survey_yr	2.82	2.31	3.29
hrrnum396:survey_yr	18.05	2.31	2.07
hrrnum397:survey_yr	0.07	-0.65	0.35
hrrnum399:survey_yr	-0.14	-0.001	0.89
hrrnum400:survey_yr			
Observations	379	379	377
Log Likelihood	-205.16	-195.84	-186.11
Akaike Inf. Crit.	506.33	507.68	508.21

*Note:* \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Table C.3: Hospitals, Both Years, Factor Models with HRR, Table 1

	<i>Dependent variable:</i>		
	ehr		
	(1)	(2)	(3)
hrrnum402:survey_yr			-17.03
hrrnum406:survey_yr			2.14
hrrnum411:survey_yr			16.72
hrrnum412:survey_yr			0.37
hrrnum413:survey_yr	16.45	14.99	16.18
hrrnum416:survey_yr	16.10	16.44	17.96
hrrnum417:survey_yr	2.28	0.83	1.46
hrrnum418:survey_yr	17.80	16.78	17.58
hrrnum420:survey_yr	1.14	1.21	1.72
urbrurcode2:survey_yr		-0.14	-0.10
urbrurcode3:survey_yr		14.60	15.63
urbrurcode4:survey_yr		-1.26	-0.43
urbrurcode5:survey_yr		-2.57**	-2.67**
urbrurcode6:survey_yr		0.33	0.65
zippovpct_lvlsBetween 10 and 20:survey_yr			1.40
zippovpct_lvlsBetween 20 and 30:survey_yr			0.65
zippovpct_lvlsBetween 30 and 40:survey_yr			1.52
zippovpct_lvlsBetween 40 and 50:survey_yr			16.26
zippovpct_lvlsMore than 50:survey_yr			2.33
Constant			
Observations	379	379	377
Log Likelihood	-205.16	-195.84	-186.11
Akaike Inf. Crit.	506.33	507.68	508.21
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01		

Table C.4: Hospitals, Both Years, Factor Models with HRR, Table 1



## Appendix D

### Hospital EHR Adoption Estimated Using Linear Probability Models

As a further test of the robustness of the coefficient estimates, the same models are estimated below as linear probability models. Under these specifications, the same coefficient estimates show similar levels of significance.

	<i>Dependent variable:</i>		
	ehr		
	(1)	(2)	(3)
Size - Medium (100-400)	0.21***	0.23***	0.23***
Size - Large (more than 400)	0.38***	0.42***	0.45***
Metro-Large Fringe		0.02	-0.01
Metro-Medium		0.06	0.02
Metro-Small		-0.11	-0.14
Metro-Micro		0.15	0.12
Metro-Noncore		0.10	0.06
Poverty pct (10-20)			0.07
Poverty (20-30)			-0.01
Poverty pct (30-40)			0.02
Poverty pct (40-50)			-0.06
Poverty pct (more than 50)			
Observations	379	379	377
Log Likelihood	-246.35	-235.65	-226.17
Akaike Inf. Crit.	504.71	503.29	504.34

*Note:*

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Table D.1: Hospitals, Both Years, LPM Factor Models, Explanatory Variables

	<i>Dependent variable:</i>		
	ehr		
	(1)	(2)	(3)
Survey Year			0.17
Size - Medium (100-400)*Year			0.01
Size - Large (more than 400)*Year			-0.12
Metro-Large Fringe*Year			-0.12
Metro-Medium*Year		-0.13	-0.09
Metro-Small*Year		-0.07	0.04
Metro-Micro*Year		-0.42**	-0.40**
Metro-Noncore*Year		0.13	0.15
Poverty pct (10-20)*Year			0.21
Poverty pct (20-30)*Year			0.07
Poverty pct (30-40)*Year			0.20
Poverty pct (40-50)*Year			0.42
Poverty pct (more than 50)*Year			0.05
Constant			
Observations	379	379	377
Log Likelihood	-246.35	-235.65	-226.17
Akaike Inf. Crit.	504.71	503.29	504.34

*Note:* \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Table D.2: Hospitals, Both Years, LPM Factor Models (cont), Year Effects

## Appendix E

### Predicting Physician EHR Adoption Using Survival Analysis

Although survival analysis using proportional hazard models was not deemed to be appropriate for the analysis of EHR adoption levels over time, due to the lack of time-varying covariates, the first several steps of the process are shown below.

#### E.1 Survival Analysis

Survival analysis (or “event history analysis”) is an alternative analytic approach that may avoid some of the pitfalls of the Bass diffusion approach discussed above. Survival analysis involves the analysis of the time to the occurrence of an event across a population and the application of this analysis to projection of future population levels. Historically, survival analysis emerged from demography and population studies, but its more general application as “event history analysis” has been applied in many different situations.

### E.1.1 Application of Survival Analysis

The practitioner data from the 2012 survey includes the year that the practitioner adopted an EHR, making the data amenable to analysis within a survival analysis framework.

### E.1.2 Non-Parametric Survival/Hazard Functions

Development of non-parametric survival and hazard functions for the adoption of EHRs using the 2012 survey data is fairly straightforward.

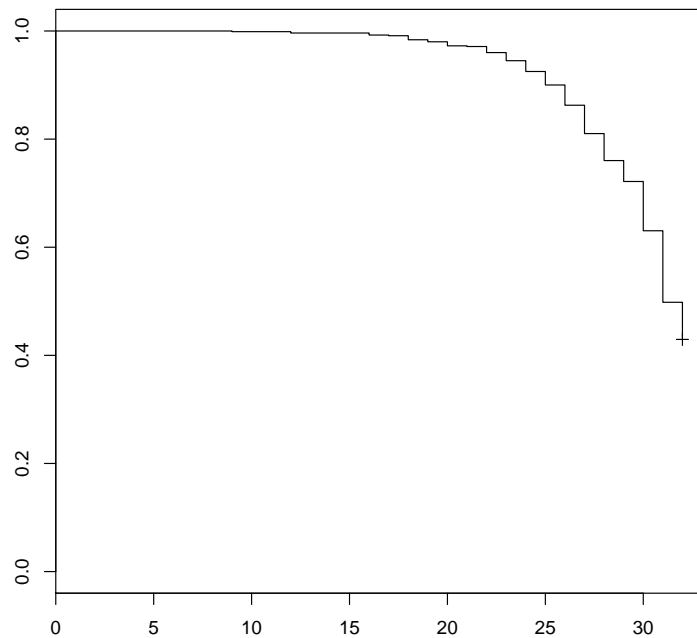


Figure E.1: Survival Curve

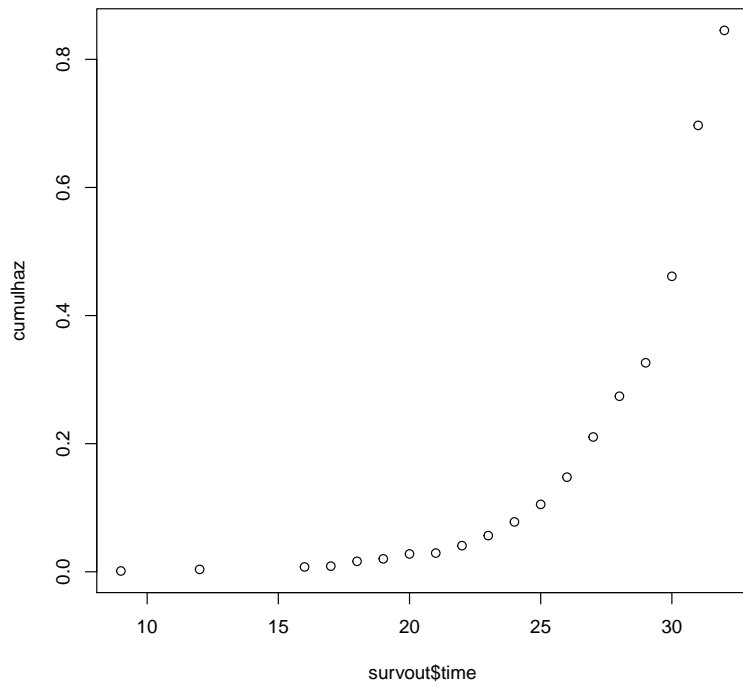


Figure E.2: Hazard Function

### **E.1.3 Proportional Hazard and Parametric Survival Functions**

Within the application of survival analysis (or ‘event history analysis’ as a more general term), one common approach is to make certain simplifying assumptions about the relative impacts of different variables or the relative impacts of certain variables across time. This approach is typified in the proportional hazard model. Another common approach is to make assumptions about the pattern of ‘survival’ or events by establishing a particular parametric form for the survival function (and, by extension, the hazard function.)

### **E.1.4 Limitations of Survival Analysis for Projection**

While it is possible to develop descriptive survival and hazard functions in the current analysis, the structure and content of the data effectively preclude meaningful extension of the survival analysis approach beyond the time period of the data. Because the purpose of this analysis is to project future EHR adoption rates, the structure and content of the available data largely preclude this goal. Or, more specifically, the lack of time-specific covariates within the data renders the potential analytic options available via the survival analysis framework no more powerful or flexible than those available through more straightforward analysis.

## Appendix F

### Bass Model Sensitivity Analysis by Effect Year

One key assumption used in constructing the confidence intervals for the Bass models in chapter 6, which enabled the formal hypothesis tests, was that the program was implemented in 2009. While the underlying legislation did pass in 2009, full implementation of the programs spanned several years. In the models below, different implementation years are tested to see if the actual measured adoption rates are greater than the 95% confidence interval calculated using a set of prior observations prior to a different implementation year.

The results using 2010 as the implementation year are very similar to those using 2009 as the implementation year. In both cases, the measured adoption levels for 2011 and 2012 are slightly above the upper 95% confidence interval, indicating statistical significance based on our formal hypothesis test.

Year	Conf. Int., Low	Measured Adoption	Conf. Int., High
2009	26.40	27.84	34.38
2010	32.48	36.95	42.78
2011	39.11	50.19	51.88
2012	46.12	57.05	60.97

Table F.1: Actual Adoption Levels and Confidence Intervals, 2008 Implementation



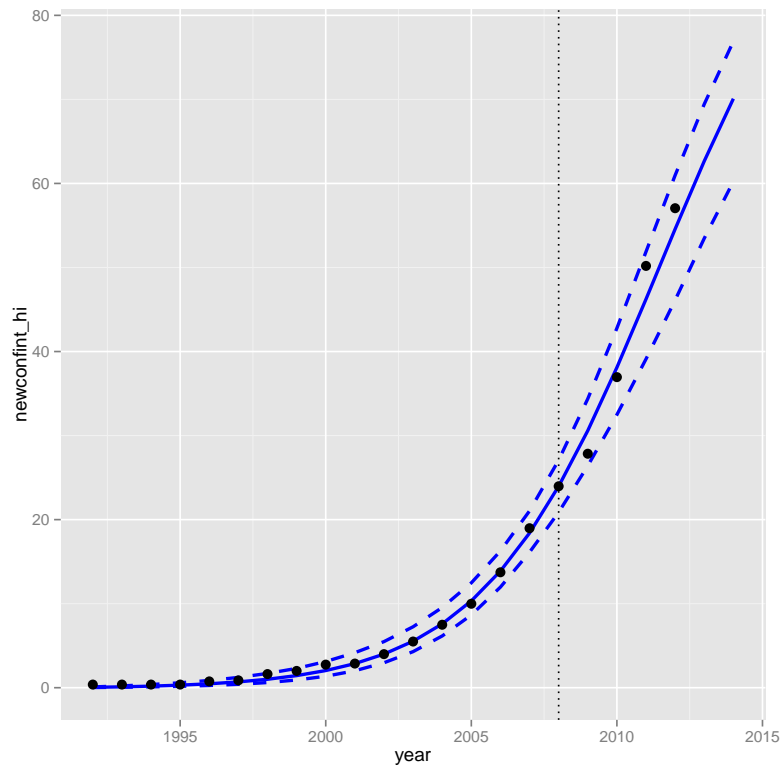


Figure F.1: Basic Confidence Interval Plots with Bootstrapping, 2008 Implementation

Year	Conf. Int., Low	Measured Adoption	Conf. Int., High
2011	39.11	50.19	51.88
2012	46.12	57.05	60.97

Table F.2: Actual Adoption Levels and Confidence Intervals, 2010 Implementation

Year	Conf. Int., Low	Measured Adoption	Conf. Int., High
2012	46.12	57.05	60.97

Table F.3: Actual Adoption Levels and Confidence Intervals, 2011 Implementation

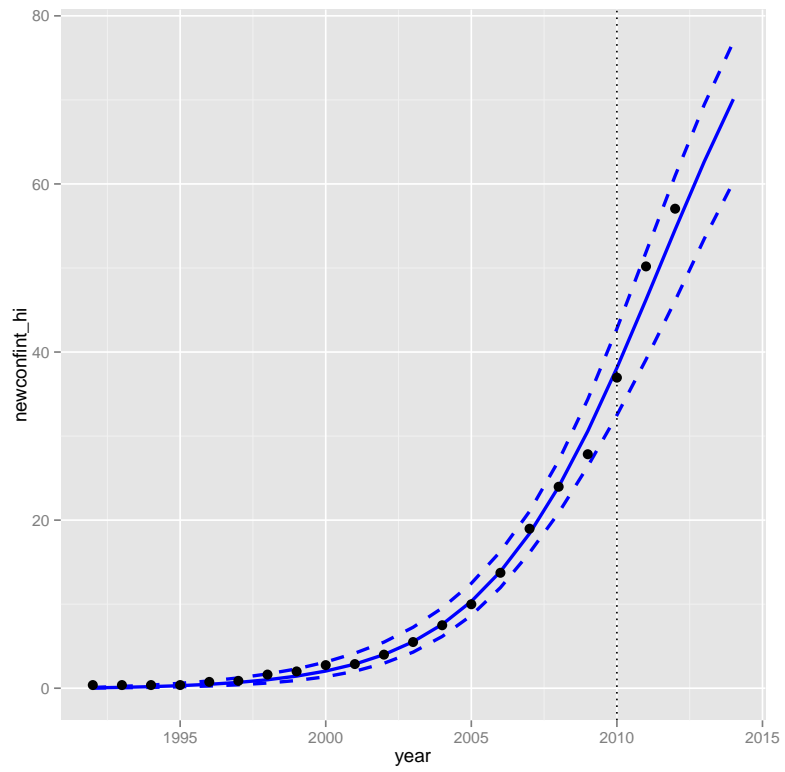


Figure F.2: Basic Confidence Interval Plots with Bootstrapping, 2010 Implementation

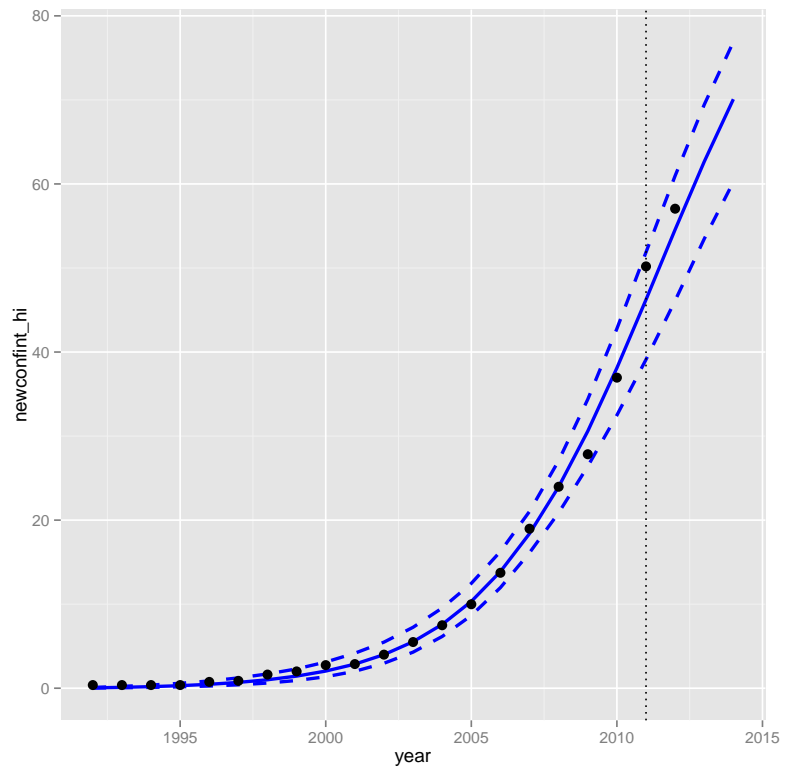


Figure F.3: Basic Confidence Interval Plots with Bootstrapping, 2011 Implementation

## Appendix G

### Robustness of Physician RD Design Models to Addition of Covariates

The following models reflect the same general approach to estimating the impact of REC programs at the 10 physician practice size taken in chapter 7, except in these models, covariates are added representing percent poverty as a discrete variable in 10 percentage point buckets, and metropolitan status, using the HRSA county-level metro-status categories previously discussed.

While the addition of the covariates changes the shapes and effect estimates of the models, it does not do so in a consistent way, nor are the effects any more significant than without the addition of the covariates. Since the coefficient estimates were not previously significant, this does not change the interpretation.

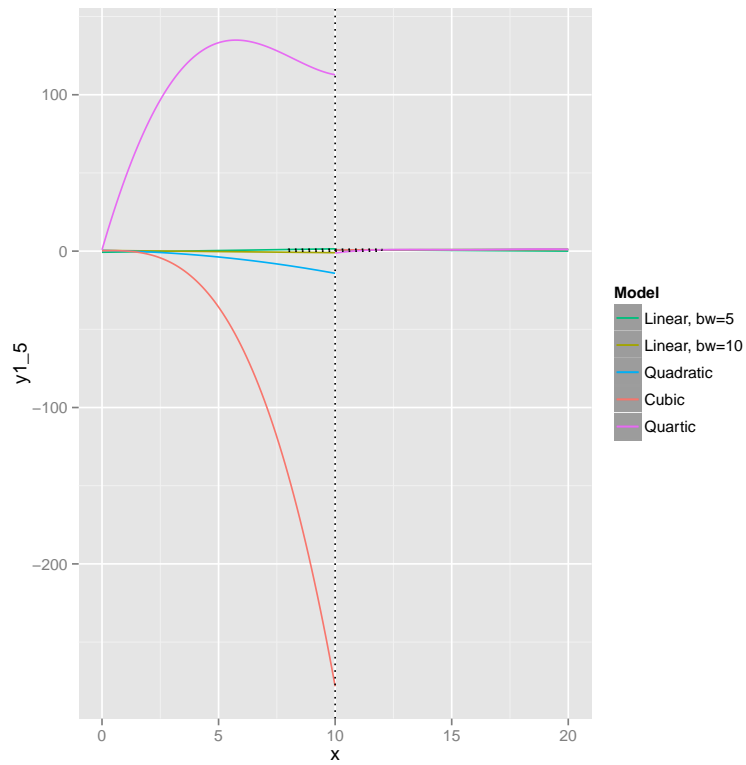


Figure G.1: REC RD Design Table - Multiple Models with Covariates

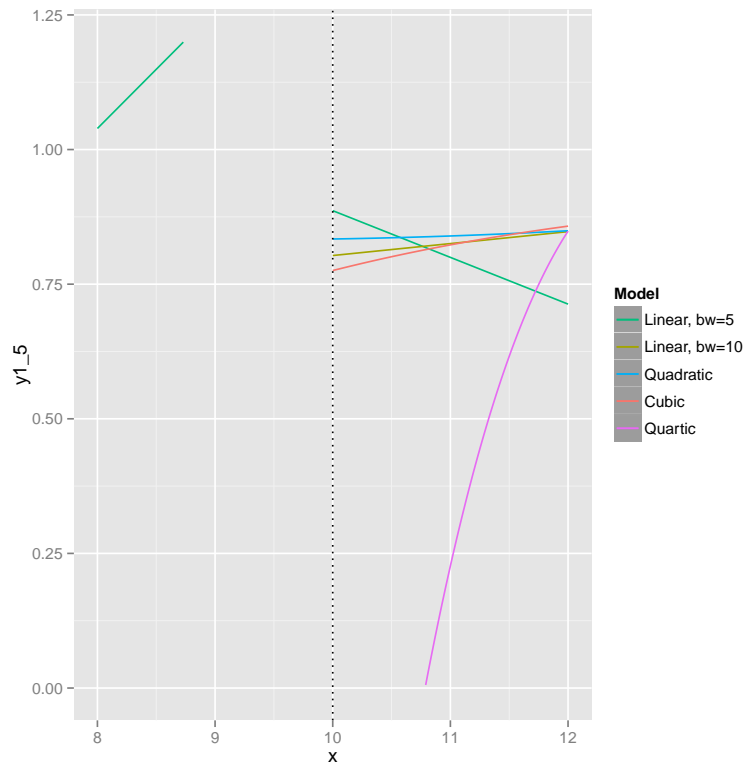


Figure G.2: REC RD design Table - Multiple Models with Covariates, Cut-off zoom

	Estimate	Pr(> t )
Local linear, bandwidth=5	-2.4720	0.7303
Local linear, bandwidth=10	-0.1579	0.8975
Quadratic	-0.5964	0.9147
Cubic	1.4724	0.9816
Quartic	195.7376	0.7590

Table G.1: REC RDD Effect Sizes at Cut-off with Covariates

## Appendix H

### Robustness of Physician RD Design Models to Different Bandwidth Calculations

The original physician RD design models did not show statistically significant effects at the cutoff. The results of additional models using different bandwidths are shown below. As in the case of the original models, none of the specifications show statistically significant estimates.

	Estimate	Pr(> t )
Local linear, bandwidth=3	1.5582	0.8162
Local linear, bandwidth=5	0.1963	0.9168
Local linear, bandwidth=10	0.0288	0.9741
Quadratic	2.2449	0.6852
Quadratic, bandwidth=5	17.1461	0.7217
Quadratic, bandwidth=10	2.2449	0.6852
Cubic	18.0058	0.7764
Cubic, bandwidth=5	8.5942	0.9424
Cubic, bandwidth=10	18.0058	0.7764
Quartic	129.5635	0.8381
Quartic, bandwidth=5	805.8409	0.7078
Quartic, bandwidth=10	129.5635	0.8381

Table H.1: Expanded REC RDD Effect Sizes at Cut-off



## Appendix I

### Robustness of Hospital RD Design Models to Addition of Covariates

The following models reflect the same general approach to estimating the impact of the Medicaid EHR incentive program at the 10 percent Medicaid caseload level taken in chapter 7, except in these models, covariates are added representing percent poverty as a discrete variable in 10 percentage point buckets, and metropolitan status, using the HRSA county-level metro-status categories previously discussed.

As in the case of the physician models with added covariates, the addition of the covariates changes the shapes and effect estimates of the models, but it does not do so in a consistent way, nor are the effects any more significant than without the addition of the covariates. Since the coefficient estimates were not previously significant, this does not change the interpretation.

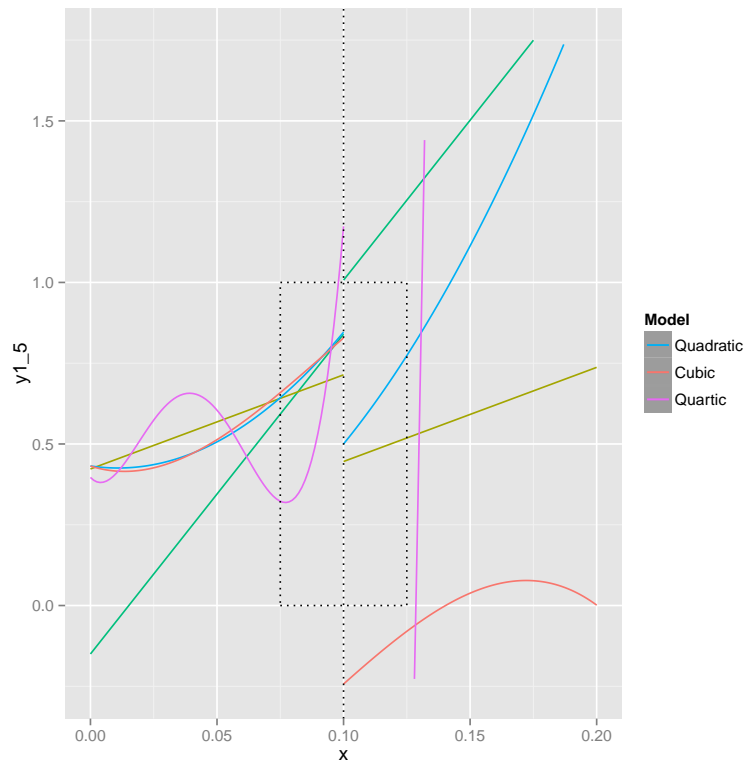


Figure I.1: MCD RD design Table with Covariates - Multiple Models

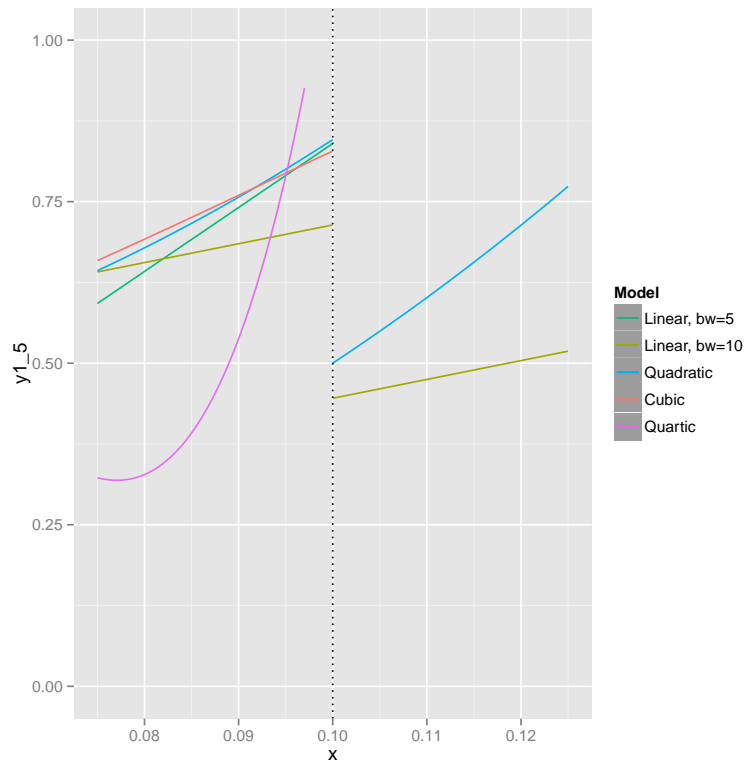


Figure I.2: MCD RD design Table with Covariates - Multiple Models, Zoom

	Estimate	Pr(> t )
Local linear, bandwidth=5	0.1671	0.9077
Local linear, bandwidth=10	-0.2682	0.6160
Quadratic	-0.3656	0.6843
Cubic	-1.0904	0.6948
Quartic	-7.5529	0.4000

Table I.1: Medicaid RDD Effect Sizes at Cut-off with Covariates

## Appendix J

### Robustness of Hospital RD Design Models to Different Bandwidth Calculations

The original hospital RD design models did not show statistically significant effects at the cutoff. The results of additional models using different bandwidths are shown below. As in the case of the original models, none of the specifications show statistically significant estimates.

	Estimate	Pr(> t )
Local linear, bandwidth=1	0.7096	0.9576
Local linear, bandwidth=3	0.8663	0.8892
Local linear, bandwidth=5	-0.0203	0.9875
Local linear, bandwidth=10	-0.3404	0.5240
Quadratic	-0.4642	0.5994
Quadratic, bandwidth=5	1.9694	0.8637
Quadratic, bandwidth=10	-1.5166	0.6293
Cubic	-1.5880	0.5616
Cubic, bandwidth=5	-123.2789	0.2102
Cubic, bandwidth=10	18.8022	0.2651
Quartic	-5.5750	0.5286
Quartic, bandwidth=5	-709.2290	0.3927
Quartic, bandwidth=10	-75.7728	0.4393

Table J.1: Medicaid RDD Effect Sizes at Cut-off

## Appendix K

### R Code Used for Statistics, Tables, and Figures

#### K.1 R Code Used for Statistics, Tables, and Figures in Chapter 3 - Data

```
# This document has been 'knit' together to include R code, R  
  output,  
# narrative text, figures, and tables.
```

```
# *R Code Block* - loadlibraries - Load non-base R libraries  
rm(list = ls())  
library(stringr)  
library(xtable)  
library(ggplot2)  
library(plyr)  
library(corrgram)  
library(stargazer)  
library(BaylorEdPsych)  
library(sas7bdat)
```

```
# *R Code Block* - setdefaults - Set R defaults and constants  
# options(width=50)  
  
figtheme ← theme(plot.title = element_text(face = "bold", size =  
  16)) + theme(axis.title.y = element_text(face = "bold",  
  size = 14), axis.text.y = element_text(size = 12, color = "  
  black"), axis.title.x = element_text(face = "bold",  
  size = 14), axis.text.x = element_text(size = 12, color = "  
  black"))  
  
options(bitmapType = "cairo")
```

```
# *R Code Block* - printtbl - Load my function to print survey  
  question
```

```

# answer options for a particular question

PrintTbl ← function(CodeBook, SurveyField) {

  c2 ← CodeBook[CodeBook$Name == SurveyField, c("Question", "
    Label", "Value")]
  c2tab ← c2[c2$Question == 0, c("Label", "Value")]
  c2lab ← c2[c2$Question == 1, "Label"]
  tabout ← xtable(c2tab, label = SurveyField, caption = c2lab,
    align = "|l|X|l|")
  print((tabout), include.rownames = FALSE, tabular.environment
    = "tabularx",
    width = "\\textwidth", type = "latex")
}

printxtab ← function(col1, col2, col1names, col2names, lbl, cap,
  aln) {
  testttab ← table(col1, col2)
  attr(testttab, "dimnames") ← list(col1names, col2names)
  xtable(testttab, type = "latex", asis = "TRUE", floating =
    FALSE, label = lbl,
    caption = cap, align = aln)
}

# *R Code Block* - loadpract11 - read pract 11 data and make pract
# 2011 data
# frame
pract2011 ← read.csv(file = str_c(dir, ":\users\\scptex\\dropbox
  \\dissertation\\data\\2011practitionersurvey_work\\
  practsurv2011.csv"),
  head = TRUE, sep = ",") #read in 2011 practitioners survey
  file
pract.sub2011 ← pract2011[pract2011$DI1 == 1 & pract2011$DI2 %in%
  c(1, 2) &
  !(pract2011$DI4 %in% c(8888, 9999)) & !(pract2011$EHR1 %in% c
  (88, 99)) &
  !(pract2011$DI3_COUNTY %in% c(88, 99)), c("resp_num", "DI2", "
  DI3_COUNTY_CODE1",
  "DI4", "EHR1", "HIE1", "EHR2.1", "EHR2.2", "EHR2.3", "EHR2.4",
  "EHR2.5",
  "EHR2.6", "EHR2.7", "EHR2.8", "EHR2.9", "EHR2.10", "EHR2.11",
  "EHR2.12",
  "EHRI1", "EHRI2", "EHR5", "DI3_ZIP")]

```

```

# *R Code Block* - loadpract12 - read pract 12 data and make pract
2012 data
# frame
pract2012 ← read.csv(file = str_c(dir, ":\users\\scptex\\dropbox
  \\dissertation\\data\\2012\\pract2012.csv"),
  head = TRUE, sep = ",")
pract.sub2012 ← pract2012[pract2012$DI1 == 1 & pract2012$DI2 %in%
  c(1, 2) &
  !(pract2012$DI4 %in% c(8888, 9999)) & !(pract2012$EHR1 %in% c
    (88, 99)) &
  !(pract2012$DI3_COUNTY %in% c(88, 99)), c("resp_num", "DI2", "
    DI3_COUNTY_CODE_1_1",
  "DI4", "EHR1", "HIE1", "EHR2_1", "EHR2_2", "EHR2_3", "EHR2_4",
  "EHR2_5",
  "EHR2_6", "EHR2_7", "EHR2_8", "EHR2_9", "EHR2_10", "EHR2_11",
  "EHR2_12",
  "EHR1", "EHR2", "EHR5", "DI3_ZIP")]

```

```

# *R Code Block* - loadcountyfiles - codeblock to load countyxref
file and
# rename county field
counties ← read.csv(file = str_c(dir, ":\users\\scptex\\dropbox
  \\dissertation\\data\\2011practitionersurvey_work\\
  countyxref.csv"),
  head = TRUE, sep = ",") #read in counties file
# names(counties)[2]='COUNTYCODE' #rename county number field in
counties
counties$COUNTYCODE ← str_pad(as.character(counties$COUNTYCODE),
  3, side = "left",
  pad = "0")
counties$POPDENSITY ← as.numeric(gsub(",", "", as.character(
  counties$Population)))/as.numeric(gsub(",", "",
  "", as.character(counties$Size)))
countyincome ← read.csv(file = str_c(dir, ":\users\\scptex\\
  dropbox\\dissertation\\data\\est10ALL.csv"),
  head = TRUE, sep = ",", skip = 2) #read in county income file
#load countyincome file, limit it to TX counties and
#relevant columns, and format county FIPS field for
#matching to other files
countyincome.tx ← countyincome[countyincome$Postal == "TX", c("
  County.FIPS",
  "Name", "Poverty.Estimate.All.Ages", "Poverty.Percent.All.Ages
  ", "Poverty.Estimate.Under.Age.18",
  "Poverty.Percent.Under.Age.18", "Poverty.Estimate.Ages.5.17",
  "Poverty.Percent.Ages.5.17",

```



```

      "Median.Household.Income", "Poverty.Estimate.Ages.0.4", "
      Poverty.Percent.Ages.0.4") ]
countyincome.tx$County.FIPS ← str_pad(as.character(
  countyincome.tx$County.FIPS),
  3, side = "left", pad = "0")
names(countyincome.tx)[which(names(countyincome.tx) == "
  County.FIPS")] ← "COUNTYCODE"
names(countyincome.tx)[which(names(countyincome.tx) == "
  Median.Household.Income")] ← "MEDIANINCOME"
names(countyincome.tx)[which(names(countyincome.tx) == "
  Poverty.Percent.All.Ages")] ← "POVERTYPCT"
countyincome.tx$MEDIANINCOME ← as.numeric(gsub(",", "", " ",
  as.character(countyincome.tx$MEDIANINCOME)))
countyincome.tx$POVERTYPCT ← as.numeric(gsub(",", "", " ",
  as.character(countyincome.tx$POVERTYPCT)))
UrbanRural ← read.sas7bdat(str_c(dir, ":\ users\ scptex\ dropbox\
  \ dissertation\ data\ NCHSUrbanRural\
  NCHSUrbruralcodes.sas7bdat"))
UrbanRural ← UrbanRural[which(UrbanRural$ST_ABBREV == "TX"), ]
UrbanRural$CO_FIPS ← str_pad(as.character(UrbanRural$CO_FIPS), 3,
  side = "left",
  pad = "0")
urbrurxref ← as.data.frame(cbind(c(1, 2, 3, 4, 5, 6), c("Large
  central metro",
  "Large fringe metro", "Medium metro", "Small metro", "
  Micropolitan", "Noncore")))
names(urbrurxref) ← c("CODE2006", "MetroStatus")
counties ← merge(counties, countyincome.tx, by = "COUNTYCODE")
counties ← merge(counties, UrbanRural, by.x = "COUNTYCODE", by.y
  = "CO_FIPS")

```

```

# *R Code Block* - loadzipfile - load file with zipcode level
# poverty level
# stats (ACS)
zippov ← read.csv(file = str_c(dir, ":\ users\ scptex\ dropbox\
  \ dissertation\ Data\ ZIPPOV\ zippovpct.csv"),
  head = TRUE, sep = ",") #read in zipcode povpct xref
names(zippov)[which(names(zippov) == "povpct")] ← "ZIPPOVPCT"
zippov$ZIPPOVPCT ← as.numeric(as.character(zippov$ZIPPOVPCT))

```

```

# *R Code Block* - mergepractyears - merge 2011 and 2012
# practitioner files
pract.sub2011$survey_yr ← 0
pract.sub2012$survey_yr ← 1
names(pract.sub2011)[3] ← "COUNTYCODE"

```

```

names(pract.sub2012)[3] ← "COUNTYCODE"
pract.sub ← rbind(pract.sub2011, pract.sub2012)

# *R Code Block* - mergepractcounty - codeblock to merge
# practitioner and
# countyxref files, create new fields, re-code existing fields,
# create
# meaningful column names, and reduce data frame width for
# relevant columns
pract.sub$COUNTYCODE ← str_pad(as.character(pract.sub$COUNTYCODE)
, 3, side = "left",
pad = "0")
# source(str_c(dir, ':\\users\\scptex\\Dropbox\\Dissertation\\Text
\\code\\Counties\\LoadCountyFiles.r'))
pract.sub ← merge(pract.sub, counties, by = "COUNTYCODE") #merge
pract.sub and counties on COUNTYCODE field
pract.sub$Population ← as.numeric(gsub(",", "", as.character(
pract.sub$Population)))
pract.sub$Size ← as.numeric(gsub(",", "", as.character(pract.sub$
Size))) #create new population density field
pract.sub$Population ← as.numeric(pract.sub$Population)
pract.sub$Size ← as.numeric(pract.sub$Size)
pract.sub$POPDENSITY ← pract.sub$Population/pract.sub$Size
pract.sub$DI2 ← as.integer(sub(2, 0, pract.sub$DI2)) #to sub
zeroes for 2s in coding for DI2, Medicaid =1, no=0 after sub
names(pract.sub)[which(names(pract.sub) == "DI2")] ← "MEDICAID"
#rename DI2 to MEDICAID
names(pract.sub)[which(names(pract.sub) == "DI4")] ← "
PRACTITIONERCOUNT" #rename DI4 to PRACTITIONERCOUNT
pract.sub$EHR1_orig ← pract.sub$EHR1
pract.sub$EHR1 ← sub(2, 1, pract.sub$EHR1) #code EHR1 2 to 1;
makes paperless or mostly paperless 1
pract.sub$EHR1 ← sub(3, 0, pract.sub$EHR1) #code EHR1 3, 4, and
5 to 0; make everything else 0
pract.sub$EHR1 ← sub(4, 0, pract.sub$EHR1)
pract.sub$EHR1 ← sub(5, 0, pract.sub$EHR1)
names(pract.sub)[which(names(pract.sub) == "EHR1")] ← "EHR" #
rename EHR1 to EHR
pract.sub$EHR ← as.numeric(pract.sub$EHR) #make EHR a numeric
field
pract.sub$EHR ← as.factor(pract.sub$EHR)
pract.sub$MEDICAID ← factor(pract.sub$MEDICAID) #make MEDICAID a
factor; this didn't seem to change anything
pract.sub$CODE2006 ← factor(pract.sub$CODE2006)
pract.sub ← merge(pract.sub, urbrurxref, by = "CODE2006")

```

```

pract.sub$MetroStatus ← factor(pract.sub$MetroStatus, levels(
  pract.sub$MetroStatus)[c(5,
    4, 6, 3, 2, 1)])
pract.sub$popdens_lvls ← as.ordered(cut(pract.sub$POPDENSITY, 4))
pract.sub$practcnt_lvls ← cut(pract.sub$PRACTITIONERCOUNT, breaks
  = c(0, 1,
    5, 10, 100, 1000), labels = c("Solo (1)", "Small (2-5)", "
    Medium (6-10)",
    "Large (11-100)", "Very Large (101-1000)"))
pract.sub$medinc_lvls ← cut(pract.sub$MEDIANINCOME, 4)
pract.sub$povpct_lvls ← cut(pract.sub$POVERTYPCT, breaks = c(0,
  10, 20, 30,
  40), labels = c("Less than 10", "Between 10 and 20", "Between
  20 and 30",
  "Between 30 and 40"))
HRRZipXref ← read.csv(file = str_c(dir, ":\ users\scptex\
  Dropbox\Dissertation\Data\\\ DartmouthHRRs\ZipHsaHrr11.csv"
  ),
  head = TRUE, sep = ",") #read in HRR-zip xref file
pract.sub ← merge(pract.sub, HRRZipXref, by.x = "DI3_ZIP", by.y =
  "zipcode11")

pract.sub ← merge(pract.sub, zippov, by.x = "DI3_ZIP", by.y = "
  zip")
pract.sub$zippovpct_lvls ← cut(pract.sub$ZIPPOVPCT, breaks = c(0,
  10, 20, 30,
  40, 50, 60, 70), labels = c("Less than 10", "Between 10 and 20
  ", "Between 20 and 30",
  "Between 30 and 40", "Between 40 and 50", "Between 50 and 60",
  "Between 60 and 70"))

survpractsizecounts2012 ← hist(pract.sub[which(pract.sub$
  survey_yr == 1), ]$PRACTITIONERCOUNT,
  plot = F, breaks = c(0, 3.5, 8.5, 49.5, 10000))$counts

survpractsizepcts2012 ← survpractsizecounts2012/sum(
  survpractsizecounts2012)

txphyssizes2012 ← c(0.71, 0.11, 0.12, 0.07)

# prop.test(survpractsizecounts2012, n=rep(sum(
  survpractsizecounts2012), 4), p=txphyssizes2012)
# teststat← txphyssizes2012*sum(survpractsizecounts2012)
# prop.test(teststat, n=rep(sum(survpractsizecounts2012), 4), p=
  txphyssizes2012)

```

```

comparephysprops <- data.frame(sample = factor(c(rep("HHSC
Practitioner Survey",
4), rep("Population Estimate (TMA Survey)", 4))), size =
factor(rep(c("Small (1-3)",
"Medium (4-8)", "Large (9-49)", "Very Large (50 and above)"),
2), levels = c("Small (1-3)",
"Medium (4-8)", "Large (9-49)", "Very Large (50 and above)")),
value = c(survpractsizepcts2012 ,
txphysizes2012))

```

```

ggplot(data = comparephysprops, aes(x = size, y = value, fill =
sample)) + geom_bar(stat = "identity",
position = position_dodge(), colour = "black") + xlab("\
nPhysician Practice Size") +
ylab("Percentage\n") + figtheme + theme(legend.title =
element_blank()) +
theme(axis.text.x = element_text(angle = -45, hjust = 0, vjust
= 1))
# vjust=0.5, size=16)

```

*# \*R Code Block\* - makehospfile for hospital 2012 survey*

```

hosp <- read.csv(file = str_c(dir, ":\users\scptex\Dropbox\
Dissertation\Data\2012HospSurvey\HospSurv2012.csv"),
head = TRUE, sep = ",")

hosp$EHR <- sub(2, 1, hosp$EHR)
hosp$EHR <- sub(3, 0, hosp$EHR)
hosp$EHR <- sub(4, 0, hosp$EHR)
hosp$EHR <- sub(5, 0, hosp$EHR)
hosp$EHR <- sub(6, 0, hosp$EHR)
hosp$EHR <- as.factor(hosp$EHR)

hosplink <- read.csv(file = str_c(dir, ":\users\scptex\Dropbox
\Dissertation\Data\Hosp2012FromDSHS\HS12CD.csv"),
head = TRUE, sep = ",")

hosplink2 <- read.csv(file = str_c(dir, ":\users\scptex\Dropbox
\Dissertation\Data\Hosp2012FromDSHS\HS12DIR.csv"),
head = TRUE, sep = ",")

hosp <- merge(hosp, hosplink, by = "FID")

hosp <- merge(hosp, hosplink2, by = "FID")

```

```

source(str_c(dir, ":\users\sctex\Dropbox\Dissertation\Text
  \code\counties\LoadCountyFiles.r"))

counties$Name ← toupper(counties$Name)
counties$Name ← gsub(" COUNTY", "", counties$Name)
counties$Name ← gsub(" ", "", counties$Name)
hosp$county ← gsub(" ", "", hosp$COUNTY)

hosp_comerge ← merge(hosp, counties, by.x = "COUNTY", by.y = "
  Name")

hosp_comerge$MEDIANINCOME ← as.numeric(gsub(",", "", as.character
  (hosp_comerge$MEDIANINCOME)))
hosp_comerge$POVERTYPCT ← as.numeric(gsub(",", "", as.character(
  hosp_comerge$POVERTYPCT)))

HRRZipXref ← read.csv(file = str_c(dir, ":\users\sctex\
  Dropbox\Dissertation\Data\ DartmouthHRRs\ZipHsaHrr11.csv"
  ),
  head = TRUE, sep = ",") #read in HRR-zip xref file

hosp_hrrmerge ← merge(hosp_comerge, HRRZipXref, by.x = "LOCOZIP",
  by.y = "zipcode11")

hosp_sub ← hosp_hrrmerge[, c("FID", "FacName", "EHR", "EHYear", "
  D1A1", "D1F1",
  "D1G1", "D2A11", "D2B11", "D2C11", "D2D11", "D3E1", "D5B1", "
  MSTAT", "POPDENSITY",
  "MEDIANINCOME", "POVERTYPCT", "hrrnum")]

names(hosp_sub) ← c("fid", "facname", "ehr", "ehryear", "
  totalbeds", "inpatientdays",
  "edvisits", "medicaredischarges", "medicaredays", "
  medicaiddischarges",
  "medicaiddays", "totalrevenue", "charitycare", "metrostat", "
  popdensity",
  "medianincome", "povertypct", "hrrnum")

hosp_sub$size ← cut(hosp_sub$totalbeds, labels = c("Small", "
  Medium", "Large"),
  breaks = c(0, 100, 400, 5000))

hosp_sub$metrostat ← gsub("1", "metro", hosp_sub$metrostat)
hosp_sub$metrostat ← gsub("2", "nonmetro", hosp_sub$metrostat)

```

```

hosp_sub$povpct_lvls ← cut(hosp_sub$povertypct, 4)
hosp_sub$metrostat ← factor(hosp_sub$metrostat)

hosp_sub$mcdpct ← hosp_sub$medicaiddays/hosp_sub$inpatientdays
hosp_sub$mcdehrelig ← ifelse(hosp_sub$mcdpct > 0.1, 1, 0)

hosp_sub1 ← hosp_sub[hosp_sub$ehr == 0, ]
hosp_sub2 ← hosp_sub[hosp_sub$ehr == 1 & hosp_sub$ehryear ≥
  2009, ]
hosp_sub ← rbind(hosp_sub1, hosp_sub2)

hospsurvcount ← nrow(hosp[which(hosp$survey_yr == 1), ])
hospsurvhist ← hist(hosp[which(hosp$survey_yr == 1), which(names(
  hosp) == "totalbeds")],
  breaks = c(0, 100, 400, 5000), plot = FALSE)
hospsurvpts ← hospsurvhist$counts/hospsurvcount

hosppopcount ← nrow(hosplink)
hosppophist ← hist(hosplink$D1A1, breaks = c(0, 100, 400, 5000),
  plot = FALSE)
hosppoppts ← hosppophist$counts/hosppopcount

comparehospprops ← data.frame(Sample = factor(c(rep("HHSC
  Hospital Survey",
  3), rep("Population Estimate (DSHS Survey)", 3))), size =
  factor(rep(c("Small (less than 100 beds)",
  "Medium (100–399 beds)", "Large (400 or more beds)"), 2),
  levels = c("Small (less than 100 beds)",
  "Medium (100–399 beds)", "Large (400 or more beds)")), value =
  c(hospsurvpts,
  hosppoppts))

ggplot(data = comparehospprops, aes(x = size, y = value, fill =
  Sample)) + geom_bar(stat = "identity",
  position = position_dodge(), colour = "black") + xlab("\
  nHospital Size") +
  ylab("Percentage\n") + figtheme + theme(legend.title =
  element_blank()) +
  theme(axis.text.x = element_text(angle = -45, hjust = 0, vjust
  = 1))

```

## K.2 R Code Used for Statistics, Tables, and Figures in Chapter 5 - Have patterns in EHR adoption been changing in ways consistent with the programmatic features of the federal EHR promotion programs?

```
# This document has been 'knit' together to include R code, R
  output,
# narrative text, figures, and tables.

# *R Code Block* - loadlibraries - Load non-base R libraries
rm(list = ls())
library(stringr)
library(xtable)
library(ggplot2)
library(plyr)
library(corrgram)
library(stargazer)
library(BaylorEdPsych)
library(sas7bdat)

# *R Code Block* - setdefaults - Set R defaults and constants

# options(width=50)

figtheme ← theme(plot.title = element_text(face = "bold", size =
  16)) + theme(axis.title.y = element_text(face = "bold",
  size = 14), axis.text.y = element_text(size = 12, color = "
  black"), axis.title.x = element_text(face = "bold",
  size = 14), axis.text.x = element_text(size = 12, color = "
  black"))

options(bitmapType = "cairo")

# *R Code Block* - printtbl - Load my function to print survey
  question
# answer options for a particular question

PrintTbl ← function(CodeBook, SurveyField) {

  c2 ← CodeBook[CodeBook$Name == SurveyField, c("Question", "
  Label", "Value")]
  c2tab ← c2[c2$Question == 0, c("Label", "Value")]
  c2lab ← c2[c2$Question == 1, "Label"]
  tabout ← xtable(c2tab, label = SurveyField, caption = c2lab,
  align = "|1|X|1|")
}
```

```

print((tabout), include.rownames = FALSE, tabular.environment
      = "tabularx",
      width = "\\textwidth", type = "latex")
}

printxtab ← function(col1, col2, col1names, col2names, lbl, cap,
  aln) {
  testttab ← table(col1, col2)
  attr(testttab, "dimnames") ← list(col1names, col2names)
  xtable(testttab, type = "latex", asis = "TRUE", floating =
    FALSE, label = lbl,
    caption = cap, align = aln)
}

```

```

# *R Code Block* - loadpract11 - read pract 11 data and make pract
  2011 data
# frame
pract2011 ← read.csv(file = str_c(dir, ":\ users\scptex\dropbox
  \dissertation\data\2011practitionersurvey_work\
  practsurv2011.csv"),
  head = TRUE, sep = ",") #read in 2011 practitioners survey
  file
pract.sub2011 ← pract2011[pract2011$DI1 == 1 & pract2011$DI2 %in%
  c(1, 2) &
  !(pract2011$DI4 %in% c(8888, 9999)) & !(pract2011$EHR1 %in% c
  (88, 99)) &
  !(pract2011$DI3.COUNTY %in% c(88, 99)), c("resp_num", "DI2", "
  DI3.COUNTY.CODE1",
  "DI4", "EHR1", "HIE1", "EHR2.1", "EHR2.2", "EHR2.3", "EHR2.4",
  "EHR2.5",
  "EHR2.6", "EHR2.7", "EHR2.8", "EHR2.9", "EHR2.10", "EHR2.11",
  "EHR2.12",
  "EHR11", "EHR12", "EHR5", "DI3_ZIP")]

```

```

# *R Code Block* - loadpract12 - read pract 12 data and make pract
  2012 data
# frame
pract2012 ← read.csv(file = str_c(dir, ":\ users\scptex\dropbox
  \dissertation\data\2012\pract2012.csv"),
  head = TRUE, sep = ",")
pract.sub2012 ← pract2012[pract2012$DI1 == 1 & pract2012$DI2 %in%
  c(1, 2) &
  !(pract2012$DI4 %in% c(8888, 9999)) & !(pract2012$EHR1 %in% c
  (88, 99)) &
  !(pract2012$DI3.COUNTY %in% c(88, 99)), c("resp_num", "DI2", "

```



```

DI3_COUNTY_CODE_1_1",
"DI4", "EHR1", "HIE1", "EHR2_1", "EHR2_2", "EHR2_3", "EHR2_4",
"EHR2_5",
"EHR2_6", "EHR2_7", "EHR2_8", "EHR2_9", "EHR2_10", "EHR2_11",
"EHR2_12",
"EHR1", "EHR2", "EHR5", "DI3_ZIP") ]

```

```

# *R Code Block* - loadcountyfiles - codeblock to load countyxref
file and
# rename county field
counties <- read.csv(file = str_c(dir, ":\users\\scptex\\dropbox
\\dissertation\\data\\2011practitionersurvey_work\\
countyxref.csv"),
  head = TRUE, sep = ",") #read in counties file
# names(counties)[2]='COUNTYCODE' #rename county number field in
counties
counties$COUNTYCODE <- str_pad(as.character(counties$COUNTYCODE),
  3, side = "left",
  pad = "0")
counties$POPDENSITY <- as.numeric(gsub(",", "", as.character(
counties$Population)))/as.numeric(gsub(",",
"", as.character(counties$Size)))
countyincome <- read.csv(file = str_c(dir, ":\users\\scptex\\
dropbox\\dissertation\\data\\est10ALL.csv"),
  head = TRUE, sep = ",", skip = 2) #read in county income file
#load countyincome file, limit it to TX counties and
relevant columns, and format county FIPS field for
matching to other files
countyincome.tx <- countyincome[countyincome$Postal == "TX", c("
County.FIPS",
"Name", "Poverty.Estimate.All.Ages", "Poverty.Percent.All.Ages",
"Poverty.Estimate.Under.Age.18",
"Poverty.Percent.Under.Age.18", "Poverty.Estimate.Ages.5.17",
"Poverty.Percent.Ages.5.17",
"Median.Household.Income", "Poverty.Estimate.Ages.0.4", "
Poverty.Percent.Ages.0.4")]
countyincome.tx$County.FIPS <- str_pad(as.character(
countyincome.tx$County.FIPS),
  3, side = "left", pad = "0")
names(countyincome.tx)[which(names(countyincome.tx) == "
County.FIPS")] <- "COUNTYCODE"
names(countyincome.tx)[which(names(countyincome.tx) == "
Median.Household.Income")] <- "MEDIANINCOME"
names(countyincome.tx)[which(names(countyincome.tx) == "
Poverty.Percent.All.Ages")] <- "POVERTYPCT"

```

```

countyincome.tx$MEDIANINCOME ← as.numeric(gsub(","," ",
  as.character(countyincome.tx$MEDIANINCOME)))
countyincome.tx$POVERTYPCT ← as.numeric(gsub(","," ",
  as.character(countyincome.tx$POVERTYPCT)))
UrbanRural ← read.sas7bdat(str_c(dir, ":\ users\ scptex\ dropbox
  \ dissertation \ data \ NCHSUrbanRural\
  NCHSUrbruralcodes.sas7bdat"))
UrbanRural ← UrbanRural[which(UrbanRural$ST_ABBREV == "TX"), ]
UrbanRural$CO_FIPS ← str_pad(as.character(UrbanRural$CO_FIPS), 3,
  side = "left",
  pad = "0")
urbrurxref ← as.data.frame(cbind(c(1, 2, 3, 4, 5, 6), c("Large
  central metro",
  "Large fringe metro", "Medium metro", "Small metro", "
  Micropolitan", "Noncore")))
names(urbrurxref) ← c("CODE2006", "MetroStatus")
counties ← merge(counties, countyincome.tx, by = "COUNTYCODE")
counties ← merge(counties, UrbanRural, by.x = "COUNTYCODE", by.y
  = "CO_FIPS")

```

```

# *R Code Block* - loadzipfile - load file with zipcode level
# poverty level
# stats (ACS)
zippov ← read.csv(file = str_c(dir, ":\ users\ scptex\ dropbox\
  dissertation \ Data \ ZIPPOV \ zippovpct.csv"),
  head = TRUE, sep = ",") #read in zipcode povpct xref
names(zippov)[which(names(zippov) == "povpct")] ← "ZIPPOVPCT"
zippov$ZIPPOVPCT ← as.numeric(as.character(zippov$ZIPPOVPCT))

```

```

# *R Code Block* - mergepractyears - merge 2011 and 2012
# practitioner files
pract.sub2011$survey_yr ← 0
pract.sub2012$survey_yr ← 1
names(pract.sub2011)[3] ← "COUNTYCODE"
names(pract.sub2012)[3] ← "COUNTYCODE"
pract.sub ← rbind(pract.sub2011, pract.sub2012)

```

```

# *R Code Block* - mergepractcounty - codeblock to merge
# practitioner and
# countyxref files, create new fields, re-code existing fields,
# create
# meaningful column names, and reduce data frame width for
# relevant columns
pract.sub$COUNTYCODE ← str_pad(as.character(pract.sub$COUNTYCODE)
  , 3, side = "left",

```

```

pad = "0")
# source(str_c(dir, ':\\users\\scptex\\Dropbox\\Dissertation\\Text
  \\code\\Counties\\LoadCountyFiles.r'))
pract.sub <- merge(pract.sub, counties, by = "COUNTYCODE") #merge
  pract.sub and counties on COUNTYCODE field
pract.sub$Population <- as.numeric(gsub(",", "", as.character(
  pract.sub$Population)))
pract.sub$Size <- as.numeric(gsub(",", "", as.character(pract.sub$
  Size))) #create new population density field
pract.sub$Population <- as.numeric(pract.sub$Population)
pract.sub$Size <- as.numeric(pract.sub$Size)
pract.sub$POPDENSITY <- pract.sub$Population/pract.sub$Size
pract.sub$DI2 <- as.integer(sub(2, 0, pract.sub$DI2)) #to sub
  zeroes for 2s in coding for DI2, Medicaid =1, no=0 after sub
names(pract.sub)[which(names(pract.sub) == "DI2")] <- "MEDICAID"
#rename DI2 to MEDICAID
names(pract.sub)[which(names(pract.sub) == "DI4")] <- "
  PRACTITIONERCOUNT" #rename DI4 to PRACTITIONERCOUNT
pract.sub$EHR1_orig <- pract.sub$EHR1
pract.sub$EHR1 <- sub(2, 1, pract.sub$EHR1) #code EHR1 2 to 1;
  makes paperless or mostly paperless 1
pract.sub$EHR1 <- sub(3, 0, pract.sub$EHR1) #code EHR1 3, 4, and
  5 to 0; make everything else 0
pract.sub$EHR1 <- sub(4, 0, pract.sub$EHR1)
pract.sub$EHR1 <- sub(5, 0, pract.sub$EHR1)
names(pract.sub)[which(names(pract.sub) == "EHR1")] <- "EHR" #
  rename EHR1 to EHR
pract.sub$EHR <- as.numeric(pract.sub$EHR) #make EHR a numeric
  field
pract.sub$EHR <- as.factor(pract.sub$EHR)
pract.sub$MEDICAID <- factor(pract.sub$MEDICAID) #make MEDICAID a
  factor; this didn't seem to change anything
pract.sub$CODE2006 <- factor(pract.sub$CODE2006)
pract.sub <- merge(pract.sub, urbrurxref, by = "CODE2006")
pract.sub$MetroStatus <- factor(pract.sub$MetroStatus, levels(
  pract.sub$MetroStatus)[c(5,
  4, 6, 3, 2, 1)])
pract.sub$popdens_lvls <- as.ordered(cut(pract.sub$POPDENSITY, 4))
pract.sub$practent_lvls <- cut(pract.sub$PRACTITIONERCOUNT, breaks
  = c(0, 1,
  5, 10, 100, 1000), labels = c("Solo (1)", "Small (2-5)", "
  Medium (6-10)",
  "Large (11-100)", "Very Large (101-1000)"))
pract.sub$medinc_lvls <- cut(pract.sub$MEDIANINCOME, 4)
pract.sub$povpct_lvls <- cut(pract.sub$POVERTYPCT, breaks = c(0,

```

```

10, 20, 30,
40), labels = c("Less than 10", "Between 10 and 20", "Between
20 and 30",
"Between 30 and 40"))
HRRZipXref ← read.csv(file = str_c(dir, ":\users\scptex\
Dropbox\Dissertation\Data\\\DartmouthHRRs\ZipHsaHrr11.csv"
),
head = TRUE, sep = ",") #read in HRR-zip xref file
pract.sub ← merge(pract.sub, HRRZipXref, by.x = "DI3_ZIP", by.y =
"zipcode11")

pract.sub ← merge(pract.sub, zippov, by.x = "DI3_ZIP", by.y = "
zip")
pract.sub$zippovpct_lvls ← cut(pract.sub$ZIPPOVPCT, breaks = c(0,
10, 20, 30,
40, 50, 60, 70), labels = c("Less than 10", "Between 10 and 20
", "Between 20 and 30",
"Between 30 and 40", "Between 40 and 50", "Between 50 and 60",
"Between 60 and 70"))

```

```

# *R Code Block* - practdistrotbl - create practitioner
distribution table
# and plot figure
EHRDistro ← as.data.frame(table(pract.sub$EHR1, pract.sub$
survey_yr)) #make xtab table of ehr adoption level and survey
year
ehrlvl1 ← "Full EHR, \n Paperless"
ehrlvl2 ← "Partial EHR, \n Mostly Paperless"
ehrlvl3 ← "EHR planned in \n the next year"
ehrlvl4 ← "EHR planned in \n the next two years"
ehrlvl5 ← "EHR not planned"
AdoptionLevelDescs ← c(ehrlvl1, ehrlvl2, ehrlvl3, ehrlvl4,
ehrlvl5)
EHRAdoptPcts ← round(EHRDistro$Freq/sum(EHRDistro$Freq) * 100)
EHRDistro ← cbind(EHRDistro, AdoptionLevelDescs, EHRAdoptPcts)
EHRDistro$AdoptionLevelDescs ← factor(EHRDistro$
AdoptionLevelDescs, levels(EHRDistro$AdoptionLevelDescs)[c(4,
5, 3, 2, 1)]) #re-order adoption level factor
names(EHRDistro)[which(names(EHRDistro) == "Var2")] ← "Year"
EHRDistro$Year ← gsub("0", "A", EHRDistro$Year) #rename 0 and 1
to 2011 and 2012; need to go through another character
substitution first because 2011 and 2012 both have zeroes and
ones in them
EHRDistro$Year ← gsub("1", "B", EHRDistro$Year)
EHRDistro$Year ← gsub("A", "2011", EHRDistro$Year)

```

```

EHRDistro$Year ← gsub("B", "2012", EHRDistro$Year)
EHRDistro ← ddply(EHRDistro, .(Year), transform, percent = Freq/
  sum(Freq) *
    100) #calculate percentages within each year

# *R Code Block* - plotpractdistrofig - plot practitioner EHR
  adoption
# distribution
percadopt ← ggplot(EHRDistro, aes(x = AdoptionLevelDescs, y =
  percent, fill = Year)) +
  geom_bar(stat = "identity", position = position_dodge()) +
  scale_fill_brewer(palette = "Set1") +
  xlab("Adoption Level\n") + ylab("\nPercentage") + theme(
  axis.title.y = element_text(face = "bold",
  size = 14), axis.text.y = element_text(size = 12, color = "
  black"), axis.title.x = element_text(face = "bold",
  size = 14), axis.text.x = element_text(size = 12, color = "
  black", angle = 90),
  plot.title = element_text(face = "bold", size = 16)) + ylim(0,
  50)
percadopt

# *R Code Block* - summarytable
pract.sub.summ ← pract.sub[, c("EHR", "MEDICAID", "practcnt_lvls",
  "zippovpct_lvls",
  "MetroStatus", "survey_yr")]
names(pract.sub.summ) ← c("EHR Adoption", "Medicaid Participation
  ", "Practice Size",
  "Percent Poverty by Zipcode", "County Metro Status", "Survey
  Year")
tableNominal(vars = pract.sub.summ[, c("EHR Adoption", "Medicaid
  Participation",
  "Practice Size", "Percent Poverty by Zipcode", "County Metro
  Status")],
  group = pract.sub.summ$"Survey Year", longtable = FALSE,
  cumsum = FALSE,
  cap = "Summary Table", lab = "tbl.practsumm")

# *R Code Block* - physcorrgram - physician correlogram
pract.sub.corr ← pract.sub[, c("POPENSITY", "PRACTITIONERCOUNT",
  "ZIPPOVPCT")]
corrgram(pract.sub.corr, order = NULL, lower.panel = panel.shade,
  upper.panel = panel.pts,
  text.panel = panel.txt, labels = c("Population Density", "
  Physician Group Size",

```

```

"Poverty % by Zipcode"))

# *R Code Block* - xtabsehr - physician crosstabs-EHR adoption and
  other
# characteristics
tableNominal(vars = pract.sub.summ[, c("Medicaid Participation", "
  Practice Size",
  "Percent Poverty by Zipcode", "County Metro Status")], group =
  pract.sub.summ$"EHR Adoption",
  longtable = FALSE, cumsum = FALSE, cap = "Cross-tabs by EHR
  Adoption", lab = "tbl.xtabsehr")

# *R Code Block* - xtabsmcd - physician crosstabs-Medicaid
  participation and
# other characteristics
tableNominal(vars = pract.sub.summ[, c("EHR Adoption", "Practice
  Size", "Percent Poverty by Zipcode",
  "County Metro Status")], group = pract.sub.summ$"Medicaid
  Participation",
  longtable = FALSE, cumsum = FALSE, cap = "Cross-tabs by
  Medicaid Participation",
  lab = "tbl.xtabsmcd")

# *R Code Block* - mergedlogits - run logits on both years of
  physician data
# pract.sub$EHR ← as.numeric(pract.sub$EHR) #use EHR as numeric for
  linear
# probability model
pract.sub$EHR ← as.factor(pract.sub$EHR)

firsttrymergedcont ← glm(EHR ~ MEDICAID + PRACTITIONERCOUNT +
  survey_yr + MEDICAID *
  survey_yr + PRACTITIONERCOUNT * survey_yr, data = pract.sub,
  family = "binomial")
secondtrymergedcont ← glm(EHR ~ MEDICAID + PRACTITIONERCOUNT +
  ZIPPOVPCT +
  survey_yr + MEDICAID * survey_yr + PRACTITIONERCOUNT *
  survey_yr + ZIPPOVPCT *
  survey_yr, data = pract.sub, family = "binomial")
thirdtrymergedcont ← glm(EHR ~ MEDICAID + PRACTITIONERCOUNT +
  ZIPPOVPCT + POPDENSITY +
  survey_yr + MEDICAID * survey_yr + PRACTITIONERCOUNT *
  survey_yr + ZIPPOVPCT *
  survey_yr + POPDENSITY * survey_yr, data = pract.sub, family =
  "binomial")

```

```

firsttrymergedfact <- glm(EHR ~ MEDICAID + practent_lvls +
  survey_yr + MEDICAID *
  survey_yr + practent_lvls * survey_yr, data = pract.sub,
  family = "binomial")
secondtrymergedfact <- glm(EHR ~ MEDICAID + practent_lvls +
  zippovpct_lvls +
  survey_yr + MEDICAID * survey_yr + practent_lvls * survey_yr +
  zippovpct_lvls *
  survey_yr, data = pract.sub, family = "binomial")
thirdtrymergedfact <- glm(EHR ~ MEDICAID + practent_lvls +
  zippovpct_lvls +
  MetroStatus + survey_yr + MEDICAID * survey_yr + practent_lvls
  * survey_yr +
  zippovpct_lvls * survey_yr + MetroStatus * survey_yr, data =
  pract.sub,
  family = "binomial")

```

```

# *R Code Block* - printpractlogitcontresults - print result
  tables for
# practitioner logits with continuous variables
stargazer(firsttrymergedcont, secondtrymergedcont,
  thirdtrymergedcont, title = "Physicians, Both Years,
  Continuous Models",
  no.space = TRUE, digits = 2, label = "mergedpractlogitcont1",
  align = TRUE,
  keep = c(1:13), report = "vc*", covariate.labels = c("Medicaid
  Provider",
  "Practice Size", "Poverty Pct (by zipcode)", "Population
  Density", "Survey Year",
  "Medicaid Provider*Year", "Practice Size*Year", "Poverty
  Pct (by zipcode)*Year",
  "Population Density*Year", "Constant"))

```

```

# *R Code Block* - processphyspr2scont - build physician pseudo-r2
  table
# with continuous variables
p1pr2 <- PseudoR2(firsttrymergedcont)
p2pr2 <- PseudoR2(secondtrymergedcont)
p3pr2 <- PseudoR2(thirdtrymergedcont)
combinedppr2s <- rbind(p1pr2, p2pr2, p3pr2)
combinedppr2s <- combinedppr2s[, c(1, 3, 4, 7, 9)]
rownames(combinedppr2s) <- c("Model 1", "Model 2", "Model 3")

```

```

# *R Code Block* - printphyspr2scont - print physician pseudo-r2
  table with

```

```

# continuous variables
xtable(combinedppr2s, align = "|r|r|r|r|r|r|", type = "latex",
  asis = "TRUE",
  floating = FALSE, digits = 4, label = "tbl.physpr2scont",
  caption = "Additional Summary Statistics, Continuous
  Models")

# *R Code Block* - printpractlogitfactresults - print result
# tables for
# practitioner logits with discrete variables
stargazer(firsttrymergedfact, secondtrymergedfact,
  thirdtrymergedfact, title = "Physicians, Both Years, Factor
  Models, Explanatory Variables",
  no.space = TRUE, digits = 2, label = "mergedpractlogitfact1",
  align = TRUE,
  keep = c(1:16), report = "vc*", covariate.labels = c("Medicaid
  Provider",
  "Practice Size - Small", "Practice Size - Medium", "
  Practice Size - Large",
  "Practice Size - Very Large", "Poverty Pct (10-20)", "
  Poverty Pct (20-30)",
  "Poverty Pct (30-40)", "Poverty Pct (40-50)", "Poverty Pct
  (50-60)",
  "Poverty Pct (60-70)", "Metro-Micro", "Metro-Small", "
  Metro-Medium",
  "Metro-Large Fringe", "Metro-Large Central"))

stargazer(firsttrymergedfact, secondtrymergedfact,
  thirdtrymergedfact, title = "Physicians, Both Years, Factor
  Models (cont.), Year Effects",
  no.space = TRUE, digits = 2, label = "mergedpractlogitfact2",
  align = TRUE,
  keep = c(17:33), report = "vc*", covariate.labels = c("Survey
  Year", "Medicaid Provider*Year",
  "Practice Size - Small*Year", "Practice Size - Medium*Year
  ", "Practice Size - Large*Year",
  "Practice Size - Very Large*Year", "Poverty Pct (10-20)*
  Year", "Poverty Pct (20-30)*Year",
  "Poverty Pct (30-40)*Year", "Poverty Pct (40-50)*Year", "
  Poverty Pct (50-60)*Year",
  "Poverty Pct (60-70)*Year", "Metro-Micro*Year", "
  Metro-Small*Year",
  "Metro-Medium*Year", "Metro-Large Fringe*Year", "
  Metro-Large Central*Year",
  "Constant"))

```



```

# *R Code Block* - processphyspr2sfact - build physician pseudo-r2
table
# with discrete variables
p1pr2 ← PseudoR2(firsttrymergedfact)
p2pr2 ← PseudoR2(secondtrymergedfact)
p3pr2 ← PseudoR2(thirdtrymergedfact)
combinedppr2s ← rbind(p1pr2, p2pr2, p3pr2)
combinedppr2s ← combinedppr2s[, c(1, 3, 4, 7, 9)]
rownames(combinedppr2s) ← c("Model 1", "Model 2", "Model 3")

# *R Code Block* - printphyspr2sfact - print physician pseudo-r2
table with
# continuous variables
xtable(combinedppr2s, align = "|r|r|r|r|r|r|", type = "latex",
  asis = "TRUE",
  floating = FALSE, digits = 4, label = "tbl.physpr2sfact",
  caption = "Additional Summary Statistics, Factor Models")

# *R Code Block* - makehospfile for hospital 2012 survey

hosp ← read.csv(file = str_c(dir, ":\users\scptex\Dropbox\
Dissertation\Data\2012HospSurvey\HospSurv2012.csv"),
  head = TRUE, sep = ",")

hosp$EHR ← sub(2, 1, hosp$EHR)
hosp$EHR ← sub(3, 0, hosp$EHR)
hosp$EHR ← sub(4, 0, hosp$EHR)
hosp$EHR ← sub(5, 0, hosp$EHR)
hosp$EHR ← sub(6, 0, hosp$EHR)
hosp$EHR ← as.factor(hosp$EHR)

hosplink ← read.csv(file = str_c(dir, ":\users\scptex\Dropbox
\Dissertation\Data\Hosp2012FromDSHS\HS12CD.csv"),
  head = TRUE, sep = ",")

hosplink2 ← read.csv(file = str_c(dir, ":\users\scptex\Dropbox
\Dissertation\Data\Hosp2012FromDSHS\HS12DIR.csv"),
  head = TRUE, sep = ",")

hosp ← merge(hosp, hosplink, by = "FID")

hosp ← merge(hosp, hosplink2, by = "FID")

source(str_c(dir, ":\users\scptex\Dropbox\Dissertation\Text
\code\counties\LoadCountyFiles.r"))

```

```

counties$Name ← toupper(counties$Name)
counties$Name ← gsub(" COUNTY", "", counties$Name)
counties$Name ← gsub(" ", "", counties$Name)
hosp$county ← gsub(" ", "", hosp$COUNTY)

hosp_comerge ← merge(hosp, counties, by.x = "COUNTY", by.y = "
  Name")

hosp_comerge$MEDIANINCOME ← as.numeric(gsub(",", "", as.character
  (hosp_comerge$MEDIANINCOME)))
hosp_comerge$POVERTYPCT ← as.numeric(gsub(",", "", as.character(
  hosp_comerge$POVERTYPCT)))

HRRZipXref ← read.csv(file = str_c(dir, ":\users\scptex\
  Dropbox\Dissertation\Data\\\DartmouthHRRs\ZipHsaHrr11.csv"
  ),
  head = TRUE, sep = ",") #read in HRR-zip xref file

hosp_hrrmerge ← merge(hosp_comerge, HRRZipXref, by.x = "LOCOZIP",
  by.y = "zipcode11")

hosp_sub ← hosp_hrrmerge[, c("FID", "FacName", "EHR", "EHRYear",
  "D1A1", "D1F1",
  "D1G1", "D2A11", "D2B11", "D2C11", "D2D11", "D3E1", "D5B1", "
  MSTAT", "POPDENSITY",
  "MEDIANINCOME", "POVERTYPCT", "hrrnum")]

names(hosp_sub) ← c("fid", "facname", "ehr", "ehryear", "
  totalbeds", "inpatientdays",
  "edvisits", "medicaredischarges", "medicaredays", "
  medicaiddischarges",
  "medicaiddays", "totalrevenue", "charitycare", "metrostat", "
  popdensity",
  "medianincome", "povertypct", "hrrnum")

hosp_sub$size ← cut(hosp_sub$totalbeds, labels = c("Small", "
  Medium", "Large"),
  breaks = c(0, 100, 400, 5000))

hosp_sub$metrostat ← gsub("1", "metro", hosp_sub$metrostat)
hosp_sub$metrostat ← gsub("2", "nonmetro", hosp_sub$metrostat)
hosp_sub$povpct_lvls ← cut(hosp_sub$povertypct, 4)
hosp_sub$metrostat ← factor(hosp_sub$metrostat)

```

```

hosp_sub$mcdpct ← hosp_sub$medicaiddays/hosp_sub$inpatientdays
hosp_sub$mcdehrelig ← ifelse(hosp_sub$mcdpct > 0.1, 1, 0)

hosp_sub1 ← hosp_sub[hosp_sub$ehr == 0, ]
hosp_sub2 ← hosp_sub[hosp_sub$ehr == 1 & hosp_sub$ehryear ≥
  2009, ]
hosp_sub ← rbind(hosp_sub1, hosp_sub2)

# *R Code Block* - hospsummstats_contvars
hosp.print.cont ← hosp[, c("totalbeds", "totalrevenue", "
  popdensity", "zippovpct",
  "charityratio")]
hosp.print.cont$totalrevenue ← round(hosp.print.cont$totalrevenue
)
names(hosp.print.cont) ← c("Total Beds", "Total Revenue (millions
)", "Population Density",
  "Poverty % by Zip", "Charity Ratio")
stargazer(hosp.print.cont, summary = TRUE, label = "
tbl.hospsummstats_cont",
  title = "Hospital Data, Summary Statistics, Continuous
  Variables", omit.summary.stat = "n",
  report = "vc*")

# *R Code Block* - hcorr12
hosp2012.corr ← hosp[, c("ehr", "totalbeds", "inpatientdays", "
  edvisits", "medicareddays",
  "medicaiddays", "totalrevenue", "charitycare", "metrostat", "
  popdensity",
  "medianincome", "povertypct")]
names(hosp2012.corr) ← c("EHR", "Total Beds", "Inpatient Days", "
  ED Visits",
  "Medicare Days", "Medicaid Days", "Total Revenue (millions)",
  "Charity Care",
  "Metro Status", "Population Density", "Median Income", "
  Poverty Pct")

# *R Code Block* - hcorrgram12
corrgram(hosp2012.corr, order = NULL, lower.panel = panel.shade,
  upper.panel = panel.pts,
  text.panel = panel.txt)

# *R Code Block* - hospxtabsehr
tableNominal(vars = hosp.print.nom[, c("Hospital Size", "County
  Metro Status",
  "Percent Poverty by Zipcode")], group = hosp.print.nom$"EHR
  Adoption", longtable = FALSE,

```

```
cumsum = FALSE, cap = "Hospital Data, Cross-tabs by EHR
Adoption", lab = "tbl.hospxtabsehr")
```

```
# *R Code Block* - runhosplogits
```

```
hospfirsttry <- glm(ehr ~ totalbeds + totalrevenue + survey_yr +
  totalbeds *
  survey_yr + totalrevenue * survey_yr, data = hosp, family = "
  binomial")
hospsecondtry <- glm(ehr ~ totalbeds + totalrevenue + popdensity +
  survey_yr +
  totalbeds * survey_yr + totalrevenue * survey_yr + popdensity
  * survey_yr,
  data = hosp, family = "binomial")
hospthirdtry <- glm(ehr ~ totalbeds + totalrevenue + popdensity +
  zippovpct +
  charityratio + survey_yr + totalbeds * survey_yr +
  totalrevenue * survey_yr +
  popdensity * survey_yr + zippovpct * survey_yr + charityratio
  * survey_yr,
  data = hosp, family = "binomial")

hospnew1 <- glm(ehr ~ size + survey_yr + size * survey_yr, data =
  hosp, family = "binomial")
hospnew2 <- glm(ehr ~ size + urbrurcode + survey_yr + size *
  survey_yr + urbrurcode *
  survey_yr, data = hosp, family = "binomial")
hospnew3 <- glm(ehr ~ size + urbrurcode + zippovpct_lvls +
  survey_yr + size *
  survey_yr + urbrurcode * survey_yr + zippovpct_lvls *
  survey_yr, data = hosp,
  family = "binomial")
```

```
# *R Code Block* - printhosplogitresultscont
```

```
stargazer(hospfirsttry, hospsecondtry, hospthirdtry, title = "
Hospitals, Both Years, Continuous Models",
  no.space = TRUE, single.row = FALSE, label = "hosplogit12-exp"
  , align = TRUE,
  covariate.labels = c("Total Beds", "Total Revenue (millions)",
  "Population Density",
  "Poverty Pct", "Charity Ratio", "Survey Year", "Total Beds
  *Year", "Total Revenue (millions)*Year",
  "Population Density*Year", "Poverty Pct*Year", "Charity
  Ratio*Year"),
  report = "vc*")
```

```

# *R Code Block* - processhosppr2scont
h1pr2 ← PseudoR2(hospfirsttry)
h2pr2 ← PseudoR2(hospsecondtry)
h3pr2 ← PseudoR2(hospthirdtry)
combinedhpr2s ← rbind(h1pr2, h2pr2, h3pr2)
combinedhpr2s ← combinedhpr2s[, c(1, 3, 4, 7, 9)]
rownames(combinedhpr2s) ← c("Model 1", "Model 2", "Model 3")

# *R Code Block* - printhosppr2scont
xtable(combinedhpr2s, align = "|r|r|r|r|r|r|", type = "latex",
  asis = "TRUE",
  floating = FALSE, digits = 4, label = "tbl.hosppr2s", caption =
  = "\ "Goodness of Fit\ " Statistics , Continuous Models")

# *R Code Block* - printhosplogitresultsfact
stargazer(hospnew1, hospnew2, hospnew3, title = "Hospitals , Both
  Years , Factor Models , Explanatory Variables",
  no.space = TRUE, single.row = FALSE, digits = 2, label = "
  hosplogitnew_exp",
  keep = 1:12, align = TRUE, covariate.labels = c("Size - Medium
  (100-400)",
  "Size - Large (more than 400)", "Metro-Large Fringe", "
  Metro-Medium",
  "Metro-Small", "Metro-Micro", "Metro-Noncore", "Poverty
  pct (10-20)",
  "Poverty (20-30)", "Poverty pct (30-40)", "Poverty pct (40
  -50)", "Poverty pct (more than 50)"),
  report = "vc*")

stargazer(hospnew1, hospnew2, hospnew3, title = "Hospitals , Both
  Years , Factor Models (cont), Year Effects",
  no.space = TRUE, single.row = FALSE, digits = 2, label = "
  hosplogitnew_yreff",
  keep = 13:30, align = TRUE, covariate.labels = c("Survey Year"
  , "Size - Medium (100-400)*Year",
  "Size - Large (more than 400)*Year", "Metro-Large Fringe*
  Year", "Metro-Medium*Year",
  "Metro-Small*Year", "Metro-Micro*Year", "Metro-Noncore*
  Year", "Poverty pct (10-20)*Year",
  "Poverty pct (20-30)*Year", "Poverty pct (30-40)*Year", "
  Poverty pct (40-50)*Year",
  "Poverty pct (more than 50)*Year", "Constant"), report = "
  vc*")

# *R Code Block* - processhosppr2sfact

```

```

h1pr22 ← PseudoR2(hospnew1)
h2pr22 ← PseudoR2(hospnew2)
h3pr22 ← PseudoR2(hospnew3)
combinedhpr2s2 ← rbind(h1pr22, h2pr22, h3pr22)
combinedhpr2s2 ← combinedhpr2s2[, c(1, 3, 4, 7, 9)]
rownames(combinedhpr2s2) ← c("Model 1", "Model 2", "Model 3")

# *R Code Block* - printhosp2sfact
xtable(combinedhpr2s2, align = "|r|r|r|r|r|r|", type = "latex",
  asis = "TRUE",
  floating = FALSE, digits = 4, label = "tbl.hosp2s2", caption
  = "\ Goodness of Fit Statistics – Factor Models")

```

### K.3 R Code Used for Statistics, Tables, and Figures in Chapter 6 - Have federal programs to promote the adoption of electronic health records changed the rate of adoption?

```
# This document has been 'knit' together to include R code, R  
  output,  
  # narrative text, figures, and tables.
```

```
# *R Code Block* - loadlibraries - Load non-base R libraries  
rm(list = ls())  
library(stringr)  
library(xtable)  
library(ggplot2)  
library(plyr)  
library(corrgram)  
library(stargazer)  
library(BaylorEdPsych)  
library(sas7bdat)
```

```
# *R Code Block* - setdefaults - Set R defaults and constants  
  
# options(width=50)  
  
figtheme ← theme(plot.title = element_text(face = "bold", size =  
  16)) + theme(axis.title.y = element_text(face = "bold",  
  size = 14), axis.text.y = element_text(size = 12, color = "  
  black"), axis.title.x = element_text(face = "bold",  
  size = 14), axis.text.x = element_text(size = 12, color = "  
  black"))  
  
options(bitmapType = "cairo")
```

```
# *R Code Block* - initvalues1  
  
# init vector for first 12 years of adoption level  
adopt_pct ← rep(0, 3)  
adopt_hi ← rep(0, 3)  
adopt_lo ← rep(0, 3)  
  
# populate years 10-12; the 3 years for which the 2006 paper has  
  data  
adopt_pct ← c(12.8, 14.4, 18.325)  
confint ← c(0.75, 1.27, sd(c(17.1, 17.2, 18, 21)))
```

```

adopt_hi ← adopt_pct + confint
adopt_lo ← adopt_pct - confint

# *R Code Block* - initvalues2 initialize vector holding years
yr ← c(1:40)

# init starting values of p and q to estimated values from 2006
  paper
p_start ← 0.00539624624901535
q_start ← 0.167297806617814

# *R Code Block* - bassmodelfunc bass diffusion model
predbass ← function(pfn_pandq, pfn_t) {
  (1 - exp(-(pfn_pandq[1] + pfn_pandq[2]) * pfn_t))/(1 + (
    pfn_pandq[2]/pfn_pandq[1]) *
    exp(-(pfn_pandq[1] + pfn_pandq[2]) * pfn_t)) * 100
}

# *R Code Block* - meandifffunc mean difference function to take
  the average
# diff between the observed adoption levels and the calculated
  adoption
# levels
meandiff ← function(mfn_pandq, mfn_func, mfn_obs, mfn_t,
  mfn_range) {
  mean(abs(mfn_obs - mfn_func(mfn_pandq, mfn_t)[mfn_range[1]:
    mfn_range[2]]))
}

# *R Code Block* - initpandq init p and q to close to their values
  in the
# spreadsheet
pandqout ← c(0.02, 0.05)

# *R Code Block* - runopt run optimization
bassopt ← optim(par = pandqout, fn = meandiff, mfn_func =
  predbass, mfn_obs = adopt_pct,
  mfn_t = yr, mfn_range = c(10, 12))

bassopt_hi ← optim(par = pandqout, fn = meandiff, mfn_func =
  predbass, mfn_obs = adopt_hi,
  mfn_t = yr, mfn_range = c(10, 12))

bassopt_lo ← optim(par = pandqout, fn = meandiff, mfn_func =
  predbass, mfn_obs = adopt_lo,
  mfn_t = yr, mfn_range = c(10, 12))

```



```

# *R Code Block* - calcbassresid06 display the individual diffs
# for the
# several years this is mostly to check some intermediate values
# against the
# 2006 spreadsheet this one should match and it does
bassresid06 ← adopt_pct - predbass(c(p_start, q_start), yr)
[10:12]

# *R Code Block* - calcoptresid14 display diffs between observed
# adoption
# levels and predicted adoption levels based on optimized p and q
bassresid14 ← adopt_pct - predbass(bassopt$par, yr)[10:12]

# *R Code Block* - initpredopt plot predicted adoption levels
# based on
# optimized p and q

out31 ← data.frame(yr = yr, pred = predbass(bassopt$par, yr), cat
= "best")
out32 ← data.frame(yr = yr, pred = predbass(bassopt_hi$par, yr),
cat = "hi")
out33 ← data.frame(yr = yr, pred = predbass(bassopt_lo$par, yr),
cat = "lo")
out3 ← rbind(out31, out32, out33)

out3$yr ← as.integer(out3$yr)
out3$pred ← as.numeric(out3$pred)

# *R Code Block* - plotpredopt1
ggplot(data = out3, aes(x = yr, y = pred, group = cat)) +
geom_point()
# need to add better axis labels, re-label x axis to show actual
# years.
# maybe add legend

# *R Code Block* - dispcompresids display the comparative
# residuals from the
# 06 study and my 14 analysis
compresids ← cbind(bassresid06, bassresid14)
colnames(compresids) ← c("Residuals (Ford, 2006)", "Residuals (
Palmer, 2014)")
rownames(compresids) ← c("2001", "2002", "2003")

xtable(compresids, caption = "Comparative Residuals", label = "
compresids",

```

```

align = "|c|c|c|", digits = -4, type = "latex", asis = "TRUE",
floating = FALSE)

# *R Code Block* - storeoptpars14
bassopt_out ← rbind(bassopt_hi$par, bassopt$par, bassopt_lo$par)

# *R Code Block* - calcpredadopt these are the predicted adoption
levels for
# 2014 under the different adoption scenarios; 1991==yr=0
lvl14opt_best ← predbass(bassopt$par, 23)
lvl14opt_hi ← predbass(bassopt_hi$par, 23)
lvl14opt_lo ← predbass(bassopt_lo$par, 23)
lvl14opt_all ← c(lvl14opt_hi, lvl14opt_best, lvl14opt_lo)

# *R Code Block* - tippingfunc function to calculate the tipping
point in
# the adoption curve
tipping ← function(tip_pandq) {
  round(log(tip_pandq[2]/tip_pandq[1])/(tip_pandq[1] + tip_pandq
[2]) + 1991)
}

# *R Code Block* - calctip tipping points for the several
scenarios
tip14opt_best ← tipping(bassopt$par)
tip14opt_hi ← tipping(bassopt_hi$par)
tip14opt_lo ← tipping(bassopt_lo$par)
tip14opt_all ← c(tip14opt_hi, tip14opt_best, tip14opt_lo)

# *R Code Block* - buildopttbl
opttbl ← cbind(bassopt_out, tip14opt_all, lvl14opt_all)

# *R Code Block* - disptable1
disptable1 ← cbind(c("Optimistic", "Best Estimate", "Conservative
"), c(0.0047,
0.0054, 0.0053), c(0.2001, 0.1673, 0.1544), c(2009, 2011,
2012), c("71.61%",
"61.93%", "56.20%"))

colnames(disptable1) ← c("Scenarios", "p", "q", "Tipping \nPoint",
, "2014 Adoption \nPercentage")

disptablx ← xtable(disptable1, caption = "Estimates from Ford et
al (2006)",
label = "disptabl", align = "|c|c|c|c|c|c|", type = "latex",
asis = "TRUE",

```

```

floating = FALSE)

print(disptab1x, include.rownames = FALSE)

# stargazer(as.data.frame(disptable1), summary=FALSE)

# *R Code Block* - disptable2
disptable2 ← cbind.data.frame(c("Optimistic", "Best Estimate", "
  Conservative"),
  opttbl)

colnames(disptable2) ← c("Scenarios", "p", "q", "Tipping \nPoint"
  , "2014 Adoption \nPercentage")

disptab2x ← xtable(disptable2, caption = "Estimates from Palmer
  (2014)", label = "disptab2",
  align = "|c|c|c|c|c|c|", digits = c(0, 0, 4, 4, 0, 2), type =
  "latex", asis = "TRUE",
  floating = FALSE)

print(disptab2x, include.rownames = FALSE)

# *R Code Block* - printlvlperms
lvlperm ← cbind.data.frame(c("2001", "2002", "2003"), adopt_pct,
  confint, adopt_pct +
  confint, adopt_pct - confint)
colnames(lvlperm) ← c("Year", "Measured Adoption Level", "Std Err
  /Dev", "Optimistic",
  "Conservative")

displvlpermx ← xtable(lvlperm, caption = "Adoption Levels and
  Standard Errors/Deviations",
  label = "displvl", align = "|c|c|c|c|c|c|", digits = c(0, 0,
  2, 2, 2, 2),
  type = "latex", asis = "TRUE", floating = FALSE)

print(displvlpermx, include.rownames = FALSE)

# *R Code Block* - buildperm testing stacking up the observed
  adoption
# levels for matrix processing
adopt_stack ← rbind(adopt_lo, adopt_pct, adopt_hi)

# construct permutations matrix

```

```

permmat ← cbind(c(rep(1, 4), rep(-1, 4)), rep(c(rep(1, 2), rep(-1, 2)), 2),
               rep(c(1, -1), 4))

# make matrix of all permutations of extrema
maxperms ← t(t(permmat) * confint)

# create data frame of permutations of extrema
extrema ← t(adopt_pct + t(maxperms))

# init data frame for output of looping extrema bass function
out ← data.frame(matrix(0, nrow = 8, ncol = 2))

# generate bass parameters for all extrema and store them in out
for (i in 1:8) {
  bassopt_ext ← optim(par = pandqout, fn = meandiff, mfn_func =
    predbass,
    mfn_obs = extrema[i, ], mfn_t = yr, mfn_range = c(10, 12))
  # print(bassopt_ext)
  out[i, ] ← bassopt_ext$par
}

# predict adoption levels in 2014 and tipping points for all
# permutations

# *R Code Block* - printperms
perms ← cbind.data.frame(extrema, out, predbass(out, 23), tipping
  (out))

colnames(perms) ← c("2001", "2002", "2003", "p", "q", "2014 Level
  ", "Tipping Point")

perms[, 4] ← as.character(round(perms[, 4], 4))
perms[, 4][perms[, 4] == max(perms[, 4])] ← str_c(max(perms[, 4]),
  ", "*"")
perms[, 4][perms[, 4] == min(perms[, 4])] ← str_c(min(perms[, 4]),
  ", "+"")

perms[, 5] ← as.character(round(perms[, 5], 4))
perms[, 5][perms[, 5] == max(perms[, 5])] ← str_c(max(perms[, 5]),
  ", "*"")
perms[, 5][perms[, 5] == min(perms[, 5])] ← str_c(min(perms[, 5]),
  ", "+"")

perms[, 6] ← as.character(round(perms[, 6], 2))

```

```

perms[, 6][perms[, 6] == max(perms[, 6])] <- str_c(max(perms[, 6]),
  "*", ")
perms[, 6][perms[, 6] == min(perms[, 6])] <- str_c(min(perms[, 6]),
  "+")

perms[, 7] <- as.character(round(perms[, 7]))
perms[, 7][perms[, 7] == max(perms[, 7])] <- str_c(max(perms[, 7]),
  "*", ")
perms[, 7][perms[, 7] == min(perms[, 7])] <- str_c(min(perms[, 7]),
  "+")

disppermx <- xtable(perms, caption = "Extrema Permutations", label =
  "perms",
  align = "|c|c|c|c|c|c|c|c|c|", digits = c(0, 2, 2, 2, 2, 2, 2,
    0), type = "latex",
  asis = "TRUE", floating = FALSE)

print(disppermx, include.rownames = FALSE)

```

```

# *R Code Block* - montecarlo
storeresults <- data.frame(n = 0, maxp = 0, minp = 0, maxq = 0,
  minq = 0, max2014 = 0,
  min2014 = 0, maxtip = 0, mintip = 0)

# generate bass parameters for all MC values and store them in
mcout
if (!file.exists("C:\\Users\\scptex\\Dropbox\\Dissertation\\Text\\
  Section2_Bass\\mcout.csv")) {
  k <- 0
  paramlist <- rep(1:100) * 10
  for (j in paramlist) {
    k <- k + 1

    mc.t1 <- runif(j, adopt_lo[1], adopt_hi[1])
    mc.t2 <- runif(j, adopt_lo[2], adopt_hi[2])
    mc.t3 <- runif(j, adopt_lo[3], adopt_hi[3])

    mc <- cbind(mc.t1, mc.t2, mc.t3)

# init MC data frame
    mcout <- data.frame(matrix(0, nrow = 1, ncol = 2))
    for (i in 1:j) {
      bassopt_ext <- optim(par = pandqout, fn = meandiff,
        mfn_func = predbass,
        mfn_obs = mc[i, ], mfn_t = yr, mfn_range = c(10,

```

```

        12))
        # print(bassopt_ext)
        mcout[i, ] ← bassopt_ext$par
    }
    n ← j
    minp ← min(mcout[1])
    maxp ← max(mcout[1])

    minq ← min(mcout[2])
    maxq ← max(mcout[2])

    max2014 ← max(predbass(mcout, 23))
    min2014 ← min(predbass(mcout, 23))

    maxtip ← max(tipping(mcout))
    mintip ← min(tipping(mcout))

    storeresults[k, ] ← c(n, maxp, minp, maxq, minq, max2014,
        min2014,
        maxtip, mintip)
}
}

```

```

# *R Code Block* - dispmcresults
if (!file.exists("C:\\Users\\scptex\\Dropbox\\Dissertation\\Text\\
Section2_Bass\\mcout.csv")) {
    predbass(mcout, 23)
    tipping(mcout)

    max(predbass(mcout, 23))
    min(predbass(mcout, 23))

    max(tipping(mcout))
    min(tipping(mcout))

    min(mcout[1])
    max(mcout[1])

    min(mcout[2])
    max(mcout[2])
}

```

```

# *R Code Block* - writemcout
if (!file.exists("C:\\Users\\scptex\\Dropbox\\Dissertation\\Text\\

```

```

Section2_Bass\\mcout.csv")) {
  write.csv(storerresults, file = "c:\\users\\scptex\\dropbox\\
    dissertation\\text\\section2_bass\\mcout.csv")
}

```

```

# *R Code Block* - readmcout
if (file.exists("C:\\Users\\scptex\\Dropbox\\Dissertation\\Text\\
Section2_Bass\\mcout.csv")) storerresults <- read.csv(file = "c
:\\users\\scptex\\dropbox\\dissertation\\text\\section2_bass\\
mcout.csv")

```

```

# *R Code Block* - plot_mcout_minp
plotmcout_minp <- ggplot(storerresults) + aes(x = n, y = minp) +
  geom_point() +
  geom_hline(yintercept = min(storerresults$minp)) + geom_smooth
  () + xlab("Iterations") +
  ylab("Minimum p")

```

```
plotmcout_minp
```

```

# *R Code Block* - plot_mcout_maxq
plotmcout_maxq <- ggplot(storerresults) + aes(x = n, y = maxq) +
  geom_point() +
  geom_hline(yintercept = max(storerresults$maxq)) + geom_smooth
  () + xlab("Iterations") +
  ylab("Maximum q")

```

```
plotmcout_maxq
```

```

# *R Code Block* - loadpract2012
source(str_c(dir, ":\users\\scptex\\Dropbox\\Dissertation\\Text
\\code\\logit_practfile\\Pract2012DataLoad.r"))

```

```

# *R Code Block* - buildayd
buildadoptyrdistro <- function(pd) {

  adoptyrdistro <- data.frame(1989:2012)
  names(adoptyrdistro) <- "Year"

  tmpadoptyrdistro <- as.data.frame(table(pd$EHR5[!(pd$EHR5 %in%
    c(88, 99))]))
  # names(b[1]) <- 'Year'

  adoptyrdistro <- merge(adoptyrdistro, tmpadoptyrdistro, by.x =
    "Year", by.y = "Var1",

```

```

    all = TRUE)

    adoptyrdistro[is.na(adoptyrdistro)] ← 0

    cumullvl ← rep(0, nrow(adoptyrdistro))
    adoptyrdistro ← cbind(adoptyrdistro, cumullvl)

    numrows ← NROW(pd$EHR5[!(pd$EHR5 %in% c(88, 99))])

    for (i in 1:nrow(adoptyrdistro)) {
      adoptyrdistro$cumullvl[i] ← sum(adoptyrdistro$Freq[1:i])
    }

    cumulpct ← rep(0, nrow(adoptyrdistro))
    adoptyrdistro ← cbind(adoptyrdistro, cumulpct)

    # adjust for part-year 2012
    jun21 ← 173
    oct12 ← 286
    midsurv ← (jun21 + oct12)/2
    len12 ← 366

    portionleft ← (len12 - midsurv)/len12
    nextyrcount ← count(pd$EHR1[which(pd$EHR1 == 3)])$freq
    alloc ← portionleft * nextyrcount

    adoptyrdistro[which(adoptyrdistro$Year == 2012), ]$Freq ←
      adoptyrdistro[which(adoptyrdistro$Year ==
        2012), ]$Freq + round(alloc)

    adoptyrdistro$cumulpct ← adoptyrdistro$cumullvl/numrows * 100

    return(adoptyrdistro)
  }

adoptyrdistro ← buildadoptyrdistro(pract.sub)

# *R Code Block* - loadcalcbassfunction
calcbass ← function(practdata, bassfunc, difffunc, yrlist,
  initpnq, aydfunc) {

  ad ← aydfunc(practdata)

  pnq_surv_pre ← optim(par = initpnq, fn = difffunc, mfn.func =

```



```

    bassfunc ,
    mfn_obs = ad$cumulpct[4:21], mfn_t = yrlist , mfn_range = c
      (1, 18))
  pnq_surv_post ← optim(par = initpnq , fn = difffunc , mfn_func
    = bassfunc ,
    mfn_obs = ad$cumulpct[4:24], mfn_t = yrlist , mfn_range = c
      (1, 21))
  return(cbind(pnq_surv_pre , pnq_surv_post))
}

```

```

# *R Code Block* - calcnewbass
a ← calcbass(pract.sub , predbass , meandiff , yr , pandqout ,
  buildadoptyrdistro)

```

```

pnq_surv_pre ← a[1, ]$pnq_surv_pre
pnq_surv_post ← a[1, ]$pnq_surv_post

```

```

# *R Code Block* - plotpracttbl
physadoptyrs ← as.data.frame(table(pract.sub$EHR5[pract.sub$EHR5
  > 1900]))
names(physadoptyrs) ← c("Year" , "Count")

ggplot(data = physadoptyrs , aes(x = Year , y = Count)) + geom_bar(
  stat = "identity" ,
  fill = "white" , color = "black")

```

```

# *R Code Block* - printcumuldisp
ggplot(data = adoptyrdistro , aes(x = Year , y = cumulpct)) +
  geom_bar(stat = "identity" ,
    fill = "white" , color = "black") + ylab("Cumulative Adoption
    Percentage")

```

```

# *R Code Block* - newbassdisptbl
newbass ← rbind.data.frame(c(pnq_surv_pre , predbass(pnq_surv_pre ,
  23) , tipping(pnq_surv_pre)) ,
  c(pnq_surv_post , predbass(pnq_surv_post , 23) , tipping(
    pnq_surv_post)))
colnames(newbass) ← c("p" , "q" , "2014 Level" , "Tipping Point")
rownames(newbass) ← c("Pre" , "Post")

dispnewbass ← xtable(newbass , caption = "Estimated Parameter
  Values" , label = "tbl.newbass" ,
  align = "|c|c|c|c|c|" , type = "latex" , asis = "TRUE" , floating
    = FALSE)
print(dispnewbass)

```

```

# *R Code Block* - bootstrap predicted adoption rates for 2011
-2014

if (!file.exists("C:\\Users\\scptex\\Dropbox\\Dissertation\\Text\\
Section3_RDD\\bootstrap.csv")) {
  predvalues <- as.data.frame(matrix(ncol = 23, nrow = 10))
  names(predvalues) <- as.character(1992:2014)

  for (j in 1:1000) {

    practrep <- pract.sub[sample(nrow(pract.sub), nrow(
      pract.sub), replace = TRUE),
      ]

    pnqprenpost <- calcbass(practrep, predbass, meandiff, yr,
      pandqout,
      buildadoptyrdistro)

    predvalues[j, ] <- predbass(pnqprenpost[1, ]$pnq_surv_pre,
      1:23)

  }
  write.csv(predvalues, file = "c:\\users\\scptex\\dropbox\\
dissertation\\text\\Section3_RDD\\bootstrap.csv")
}

if (file.exists("C:\\Users\\scptex\\Dropbox\\Dissertation\\Text\\
Section3_RDD\\bootstrap.csv")) {
  predvalues <- read.csv(file = "c:\\users\\scptex\\dropbox\\
dissertation\\text\\Section3_RDD\\bootstrap.csv",
    header = TRUE, row.names = 1)
  names(predvalues) <- 1992:2014
}

newconfint_lo <- apply(predvalues, 2, sort)[26, ]
newconfint_hi <- apply(predvalues, 2, sort)[975, ]
predpre <- as.data.frame(predbass(pnq_surv_pre, 1:23))
names(predpre) <- "Predicted"
row.names(predpre) <- 1992:2014

# *R Code Block* - basicconfintplots
ncil <- as.data.frame(newconfint_lo)
ncih <- as.data.frame(newconfint_hi)

ncih_df <- as.data.frame(ncih)

```

```

ncih_df$year ← as.numeric(row.names(ncih_df))

ncil_df ← as.data.frame(ncil)
ncil_df$year ← as.numeric(row.names(ncil_df))

predpre$year ← as.numeric(row.names(predpre))

ggplot() + geom_line(data = ncih_df, aes(x = year, y =
  newconfint_hi, group = 1),
  color = "blue", size = 1, linetype = 2) + geom_line(data =
  ncil_df, aes(x = year,
  y = newconfint_lo, group = 1), color = "blue", size = 1,
  linetype = 2) +
  geom_line(data = predpre, aes(x = year, y = Predicted, group =
  1), color = "blue",
  size = 1) + geom_point(data = adoptyrdistro[which(
  adoptyrdistro$Year %in%
  1992:2012), ], aes(x = Year, y = cumulpct), size = 3) +
  geom_vline(xintercept = 2009,
  linetype = "dotted")

```

```

# *R Code Block* - plothyptest
hyptest ← adoptyrdistro[which(adoptyrdistro$Year %in% 2010:2012),
  c("cumulpct",
  "Year")]
names(hyptest) ← c("cumulpct", "year")

hyptest ← merge(hyptest, ncih_df, by = "year")
hyptest ← merge(hyptest, ncil_df, by = "year")
disphyptest ← hyptest[, c(1, 4, 2, 3)]
row.names(disphyptest) ← NULL

names(disphyptest) ← c("Year", "Conf. Int. , Low", "Measured
  Adoption", "Conf. Int. , High")

xdisphyptest ← xtable(disphyptest, caption = "Actual Adoption
  Levels and Confidence Intervals",
  label = "tbl.hyptest", align = "|c|c|c|c|c|", type = "latex",
  asis = "TRUE",
  floating = FALSE)

print(xdisphyptest, include.row.names = FALSE)

# this is the predicted adoption level for 2014 under the adoption
  scenario;

```

```
# 1991==yr=0
```

## K.4 R Code Used for Statistics, Tables, and Figures in Chapter 7 - Have federal programs to promote the adoption of electronic health records changed the level of adoption??

```
# *R Code Block* - loadlibraries - Load non-base R libraries
rm(list = ls())
library(stringr)
library(xtable)
library(ggplot2)
library(plyr)
library(corrgram)
library(stargazer)
library(BaylorEdPsych)
library(sas7bdat)

# *R Code Block* - init
figtheme ← theme(plot.title = element_text(face = "bold", size =
  16)) + theme(axis.title.y = element_text(face = "bold",
  size = 14), axis.text.y = element_text(size = 12, color = "
  black"), axis.title.x = element_text(face = "bold",
  size = 14), axis.text.x = element_text(size = 12, color = "
  black"))

dir ← "C"

# *R Code Block* - recrddprocess
source(str_c(dir, ":\users\sptex\Dropbox\Dissertation\Text
  \code\logit_practfile\Pract2012DataLoad.r"))

pract.sub$tenorfewer ← cut(pract.sub$DI4, breaks = c(0, 11, 1000)
  , labels = c(1,
  0))
pract.sub$recelig ← pract.sub$tenorfewer == 1 & pract.sub$DI7 ==
  1

names(pract.sub)[3] ← "COUNTYCODE"
pract.sub$COUNTYCODE ← as.character(pract.sub$COUNTYCODE)
pract.sub$COUNTYCODE ← str_pad(pract.sub$COUNTYCODE, 3, side = "
  left", pad = "0")

source(str_c(dir, ":\users\sptex\Dropbox\Dissertation\Text
  \code\Counties\LoadCountyFiles.r"))
```

```

pract.sub ← merge(pract.sub, counties, by = "COUNTYCODE")
pract.sub$Population ← as.numeric(gsub(",", "", as.character(
  pract.sub$Population)))
pract.sub$Size ← as.numeric(gsub(",", "", as.character(pract.sub$
  Size)))
pract.sub$Population ← as.numeric(pract.sub$Population)
pract.sub$Size ← as.numeric(pract.sub$Size)
pract.sub$POPDENSITY ← pract.sub$Population/pract.sub$Size

# to sub zeroes for 2s in coding for DI2, Medicaid =1, no=0 after
sub
pract.sub$DI2 ← as.integer(sub(2, 0, pract.sub$DI2))

names(pract.sub)[which(names(pract.sub) == "DI2")] ← "MEDICAID"
names(pract.sub)[which(names(pract.sub) == "DI4")] ← "
  PRACTITIONERCOUNT"

pract.sub$EHR1 ← sub(2, 1, pract.sub$EHR1)
pract.sub$EHR1 ← sub(3, 0, pract.sub$EHR1)
pract.sub$EHR1 ← sub(4, 0, pract.sub$EHR1)
pract.sub$EHR1 ← sub(5, 0, pract.sub$EHR1)

pract.sub$DI7 ← sub(2, 0, pract.sub$DI7)
names(pract.sub)[which(names(pract.sub) == "DI7")] ← "PRIMARYCARE
"
pract.sub$PRIMARYCARE ← as.factor(pract.sub$PRIMARYCARE)

names(pract.sub)[which(names(pract.sub) == "EHR1")] ← "EHR"
pract.sub$EHR ← as.numeric(pract.sub$EHR)
pract.sub$EHR ← as.factor(pract.sub$EHR)

pract.sub$MEDICAID ← factor(pract.sub$MEDICAID)
pract.sub$CODE2006 ← factor(pract.sub$CODE2006)

pract.sub ← merge(pract.sub, urbrurxref, by = "CODE2006")

pract.sub$MetroStatus ← factor(pract.sub$MetroStatus, levels(
  pract.sub$MetroStatus)[c(5,
  4, 6, 3, 2, 1)])
pract.sub$popdens_lvls ← as.ordered(cut(pract.sub$POPDENSITY, 4))
pract.sub$practent_lvls ← cut(pract.sub$PRACTITIONERCOUNT, breaks
  = c(0, 1,
  5, 10, 100, 1000), labels = c("Solo", "Small", "Medium", "
  Large", "Very Large"))

```

```

pract.sub$medinc_lvls ← cut(pract.sub$MEDIANINCOME, 4)
pract.sub$povpct_lvls ← cut(pract.sub$POVERTYPCT, 4)

pract.sub ← pract.sub[!(pract.sub$PRIMARYCARE %in% c(88, 99)), ]

pract.sub1 ← pract.sub[pract.sub$EHR == 0 & pract.sub$PRIMARYCARE
  == 1, ]
pract.sub2 ← pract.sub[pract.sub$EHR == 1 & pract.sub$PRIMARYCARE
  == 1 & pract.sub$EHR5 ≥
  2009, ]
pract.sub ← rbind(pract.sub1, pract.sub2)

pract.sub$EHR ← as.numeric(as.character(pract.sub$EHR))

recrddtest ← RDestimate(EHR ~ PRACTITIONERCOUNT, data = pract.sub
  , cutpoint = 10)

# Imbens-Kalyanaraman Optimal Bandwidth Calculation
recIKbwtst ← IKbandwidth(pract.sub$PRACTITIONERCOUNT, pract.sub$
  EHR, cutpoint = 10)

recrddlogit1 ← glm(EHR ~ MEDICAID + PRACTITIONERCOUNT + recelig,
  data = pract.sub,
  family = "binomial")
recrddlogit2 ← glm(EHR ~ MEDICAID + PRACTITIONERCOUNT + recelig +
  povpct_lvls,
  data = pract.sub, family = "binomial")
recrddlogit3 ← glm(EHR ~ MEDICAID + PRACTITIONERCOUNT + recelig +
  povpct_lvls +
  MetroStatus, data = pract.sub, family = "binomial")

# *R Code Block* - recplotfittedcurve
ggplot(pract.sub[which(pract.sub$PRACTITIONERCOUNT < 30), ]) + aes
(x = PRACTITIONERCOUNT,
  y = EHR) + geom_point(position = position_jitter(width = 0.25,
  height = 0.05),
  alpha = 0.5) + geom_smooth(method = "loess") + geom_vline(
  xintercept = 10,
  linetype = "dotted")

# *R Code Block* - recbinscount

# hist(pract.sub[which(pract.sub$PRACTITIONERCOUNT<30),]$
  PRACTITIONERCOUNT,
# breaks=seq(0,20,1))

```

```

rechist ← hist(pract.sub[which(pract.sub$PRACTITIONERCOUNT < 30),
]$PRACTITIONERCOUNT,
  breaks = seq(0, 30, 1), plot = F)

recdensfunc ← as.data.frame(cbind(rechist$mids, rechist$counts))
names(recdensfunc) ← c("mids", "counts")

ggplot() + geom_histogram(data = pract.sub[which(pract.sub$
  PRACTITIONERCOUNT <
  30), ], aes(x = PRACTITIONERCOUNT), color = "black", fill = "
  white", binwidth = 1) +
  xlim(0, 30) + geom_vline(xintercept = 10, linetype = "dotted")
  + geom_smooth(data = recdensfunc,
  aes(x = mids, y = counts))

## ---- recbinsavg # *R Code Block* - recbinsavg
# tmp ← table(pract.sub$EHR, pract.sub$PRACTITIONERCOUNT)
# tmp2 ← tmp[1,] + tmp[2,] tmp3 ← tmp[2,] / tmp2
# tmp4 ← as.data.frame(tmp3[which(as.numeric(names(tmp3)) < 30)],
  names = 'avgs')
# names(tmp4) ← 'avgs' tmp4$val ← as.numeric(row.names(tmp4))
# ggplot(data = tmp4, aes(x = val, y = avgs)) + geom_bar(color = 'black',
# stat = 'identity', fill = 'white') + geom_vline(xintercept = 10,
# linetype = 'dotted')

# *R Code Block* - calcrcpolyplots

rec_lin_bw5 ← lm(EHR ~ PRACTITIONERCOUNT + recelig +
  PRACTITIONERCOUNT * recelig,
  data = pract.sub[which(pract.sub$PRACTITIONERCOUNT < 15 &
  pract.sub$PRACTITIONERCOUNT >
  5), ])

rec_lin_bw10 ← lm(EHR ~ PRACTITIONERCOUNT + recelig +
  PRACTITIONERCOUNT * recelig,
  data = pract.sub[which(pract.sub$PRACTITIONERCOUNT < 20), ])

rec_quad ← lm(EHR ~ PRACTITIONERCOUNT + I(PRACTITIONERCOUNT^2) +
  recelig +
  PRACTITIONERCOUNT * recelig + I(PRACTITIONERCOUNT^2) * recelig
  , data = pract.sub[which(pract.sub$PRACTITIONERCOUNT <
  30), ])

```



```

rec_cub ← lm(EHR ~ PRACTITIONERCOUNT + I(PRACTITIONERCOUNT^2) + I
(PRACTITIONERCOUNT^3) +
  recelig + PRACTITIONERCOUNT * recelig + I(PRACTITIONERCOUNT^2)
  * recelig +
  I(PRACTITIONERCOUNT^3) * recelig, data = pract.sub[which(
  pract.sub$PRACTITIONERCOUNT <
  30), ])

rec_quart ← lm(EHR ~ PRACTITIONERCOUNT + I(PRACTITIONERCOUNT^2) +
  I(PRACTITIONERCOUNT^3) +
  I(PRACTITIONERCOUNT^4) + recelig + PRACTITIONERCOUNT * recelig
  + I(PRACTITIONERCOUNT^2) *
  recelig + I(PRACTITIONERCOUNT^3) * recelig + I(
  PRACTITIONERCOUNT^4) * recelig,
  data = pract.sub[which(pract.sub$PRACTITIONERCOUNT < 30), ])

rec_disptbl_pt1 ← data.frame(x = seq(0, 10, by = 0.01), y1_5 = 0,
  y1_10 = 0,
  y2 = 0, y3 = 0, y4 = 0)
rec_disptbl_pt2 ← data.frame(x = seq(10, 20, by = 0.01), y1_5 =
  0, y1_10 = 0,
  y2 = 0, y3 = 0, y4 = 0)

coefs_rec_lin_bw5 ← coefficients(rec_lin_bw5)
coefs_rec_lin_bw10 ← coefficients(rec_lin_bw10)
coefs_rec_quad ← coefficients(rec_quad)
coefs_rec_cub ← coefficients(rec_cub)
coefs_rec_quart ← coefficients(rec_quart)

# calculate linear bw=5 estimates
rec_disptbl_pt1$y1_5 ← coefs_rec_lin_bw5[1] + coefs_rec_lin_bw5
[2] * rec_disptbl_pt1$x +
  (coefs_rec_lin_bw5[2] + coefs_rec_lin_bw5[4]) *
  rec_disptbl_pt1$x + coefs_rec_lin_bw5[3]
rec_disptbl_pt2$y1_5 ← coefs_rec_lin_bw5[1] + coefs_rec_lin_bw5
[2] * rec_disptbl_pt2$x

# calculate linear bw=10 estimates
rec_disptbl_pt1$y1_10 ← coefs_rec_lin_bw10[1] + (
  coefs_rec_lin_bw10[2] + coefs_rec_lin_bw10[4]) *
  rec_disptbl_pt1$x + coefs_rec_lin_bw10[3]
rec_disptbl_pt2$y1_10 ← coefs_rec_lin_bw10[1] +
  coefs_rec_lin_bw10[2] * rec_disptbl_pt2$x

# calculate quadratic estimates

```

```

rec_disptbl_pt1$y2 ← coefs_rec_quad [1] + (coefs_rec_quad [2] +
  coefs_rec_quad [5]) *
  rec_disptbl_pt1$x + (coefs_rec_quad [3] + coefs_rec_quad [6]) *
  I(rec_disptbl_pt1$x^2) +
  coefs_rec_quad [4]
rec_disptbl_pt2$y2 ← coefs_rec_quad [1] + coefs_rec_quad [2] *
  rec_disptbl_pt2$x +
  coefs_rec_quad [3] * I(rec_disptbl_pt2$x^2)

# calculate cubic estimates
rec_disptbl_pt1$y3 ← coefs_rec_cub [1] + (coefs_rec_cub [2] +
  coefs_rec_cub [6]) *
  rec_disptbl_pt1$x + (coefs_rec_cub [3] + coefs_rec_cub [7]) * I(
  rec_disptbl_pt1$x^2) +
  (coefs_rec_cub [4] + coefs_rec_cub [8]) * I(rec_disptbl_pt1$x^3)
  + coefs_rec_cub [5]
rec_disptbl_pt2$y3 ← coefs_rec_cub [1] + coefs_rec_cub [2] *
  rec_disptbl_pt2$x +
  coefs_rec_cub [3] * I(rec_disptbl_pt2$x^2) + coefs_rec_cub [4] *
  I(rec_disptbl_pt2$x^3)

# calculate quartic estimates
rec_disptbl_pt1$y4 ← coefs_rec_quart [1] + (coefs_rec_quart [2] +
  coefs_rec_quart [7]) *
  rec_disptbl_pt1$x + (coefs_rec_quart [3] + coefs_rec_quart [8])
  * I(rec_disptbl_pt1$x^2) +
  (coefs_rec_quart [4] + coefs_rec_quart [9]) * I(rec_disptbl_pt1$x
  ^3) + (coefs_rec_quart [5] +
  coefs_rec_quart [10]) * I(rec_disptbl_pt1$x^4) +
  coefs_rec_quart [6]

rec_disptbl_pt2$y4 ← coefs_rec_quart [1] + coefs_rec_quart [2] *
  rec_disptbl_pt2$x +
  coefs_rec_quart [3] * I(rec_disptbl_pt2$x^2) + coefs_rec_quart
  [4] * I(rec_disptbl_pt2$x^3) +
  coefs_rec_quart [5] * I(rec_disptbl_pt2$x^4)

# *R Code Block* - disprecpolyplot
ggplot() + geom_line(data = rec_disptbl_pt1 , aes(x = x, y = y1.5 ,
  color = "Linear , bw=5")) +
  geom_line(data = rec_disptbl_pt2 , aes(x = x, y = y1.5 , color =
  "Linear , bw=5")) +
  geom_line(data = rec_disptbl_pt1 , aes(x = x, y = y1.10 , color
  = "Linear , bw=10")) +
  geom_line(data = rec_disptbl_pt2 , aes(x = x, y = y1.10 , color

```

```

    = "Linear , bw=10")) +
geom_smooth(data = rec_disptbl_pt1 , aes(x = x, y = y2, color =
  "Quadratic"),
  method = "lm", formula = y ~ x + I(x^2)) + geom_smooth(
  data = rec_disptbl_pt2 ,
aes(x = x, y = y2, color = "Quadratic"), method = "lm",
  formula = y ~ x +
  I(x^2)) + geom_smooth(data = rec_disptbl_pt1 , aes(x = x, y
  = y3, color = "Cubic"),
method = "lm", formula = y ~ x + I(x^2) + I(x^3)) +
geom_smooth(data = rec_disptbl_pt2 ,
aes(x = x, y = y3, color = "Cubic"), method = "lm", formula =
  y ~ x + I(x^2) +
  I(x^3)) + geom_smooth(data = rec_disptbl_pt1 , aes(x = x, y
  = y4, color = "Quartic"),
method = "lm", formula = y ~ x + I(x^2) + I(x^3) + I(x^4)) +
geom_smooth(data = rec_disptbl_pt2 ,
aes(x = x, y = y4, color = "Quartic"), method = "lm", formula
  = y ~ x +
  I(x^2) + I(x^3) + I(x^4)) + geom_vline(xintercept = 10,
  linetype = "dotted") +
geom_rect(mapping = aes(xmin = 8, xmax = 12, ymin = 0, ymax =
  1.2), fill = NA,
  color = "black", linetype = "dotted") +
  scale_color_discrete(name = "Model",
breaks = c("Linear , bw=5", "Linear , bw=10", "Quadratic", "
  Cubic", "Quartic"))

```

*# \*R Code Block\* - disprecpolyplotzoom*

```

ggplot() + geom_line(data = rec_disptbl_pt1 , aes(x = x, y = y1_5,
  color = "Linear , bw=5")) +
  geom_line(data = rec_disptbl_pt2 , aes(x = x, y = y1_5, color =
  "Linear , bw=5")) +
  geom_line(data = rec_disptbl_pt1 , aes(x = x, y = y1_10, color
  = "Linear , bw=10")) +
  geom_line(data = rec_disptbl_pt2 , aes(x = x, y = y1_10, color
  = "Linear , bw=10")) +
  geom_smooth(data = rec_disptbl_pt1 , aes(x = x, y = y2, color =
  "Quadratic"),
  method = "lm", formula = y ~ x + I(x^2)) + geom_smooth(
  data = rec_disptbl_pt2 ,
aes(x = x, y = y2, color = "Quadratic"), method = "lm",
  formula = y ~ x +
  I(x^2)) + geom_smooth(data = rec_disptbl_pt1 , aes(x = x, y
  = y3, color = "Cubic"),

```

```

method = "lm", formula = y ~ x + I(x^2) + I(x^3)) +
  geom_smooth(data = rec_disptbl_pt2 ,
aes(x = x, y = y3, color = "Cubic"), method = "lm", formula =
  y ~ x + I(x^2) +
  I(x^3)) + geom_smooth(data = rec_disptbl_pt1 , aes(x = x, y
  = y4, color = "Quartic"),
method = "lm", formula = y ~ x + I(x^2) + I(x^3) + I(x^4)) +
  geom_smooth(data = rec_disptbl_pt2 ,
aes(x = x, y = y4, color = "Quartic"), method = "lm", formula
  = y ~ x +
  I(x^2) + I(x^3) + I(x^4)) + geom_vline(xintercept = 10,
  linetype = "dotted") +
xlim(8, 12) + ylim(0, 1.2) + scale_color_discrete(name = "
  Model", breaks = c("Linear , bw=5",
  "Linear , bw=10", "Quadratic", "Cubic", "Quartic"))

```

```

# *R Code Block* - recefftbl

```

```

rec_eff_bw5 <- summary(rec_lin_bw5)$coefficients[3, c(1, 4)]
rec_eff_bw10 <- summary(rec_lin_bw10)$coefficients[3, c(1, 4)]
rec_eff_quad <- summary(rec_quad)$coefficients[4, c(1, 4)]
rec_eff_cub <- summary(rec_cub)$coefficients[5, c(1, 4)]
rec_eff_quart <- summary(rec_quart)$coefficients[6, c(1, 4)]

rec_eff <- rbind(rec_eff_bw5, rec_eff_bw10, rec_eff_quad,
  rec_eff_cub, rec_eff_quart)
rownames(rec_eff) <- c("Local linear , bandwidth=5", "Local linear ,
  bandwidth=10",
  "Quadratic", "Cubic", "Quartic")

xtable(rec_eff, type = "latex", align = "|r|r|r|", asis = "TRUE",
  floating = FALSE,
  digits = 4, label = "rec_eff", caption = "REC RDD Effect Sizes
  at Cut-off")

```

```

# *R Code Block* - buildrecrddtbls

```

```

recrddtbl <- as.data.frame(predict(recrddlogit3, type = "response"
))
recrddtbl <- cbind(recrddtbl, pract.sub$PRACTITIONERCOUNT[
  as.numeric(row.names(recrddtbl))])
names(recrddtbl) <- c("predpct", "practcount")

recrddtbl2 <- as.data.frame(predict(recrddlogit3, type = "response
"))
recrddtbl2 <- cbind(recrddtbl2, pract.sub$PRACTITIONERCOUNT[

```

```

    as.numeric(row.names(recrddtbl2))])
names(recrddtbl2) ← c("predpct", "practcount")

# as above, this is using the same logit model, but for the first
# plot it
# does not include the breakpoint

# *R Code Block* - makehospfile for hospital 2012 survey

hosp ← read.csv(file = str_c(dir, ":\users\scptex\Dropbox\
  Dissertation\Data\2012HospSurvey\HospSurv2012.csv"),
  head = TRUE, sep = ",")

hosp$EHR ← sub(2, 1, hosp$EHR)
hosp$EHR ← sub(3, 0, hosp$EHR)
hosp$EHR ← sub(4, 0, hosp$EHR)
hosp$EHR ← sub(5, 0, hosp$EHR)
hosp$EHR ← sub(6, 0, hosp$EHR)
hosp$EHR ← as.factor(hosp$EHR)

hosplink ← read.csv(file = str_c(dir, ":\users\scptex\Dropbox
  \Dissertation\Data\Hosp2012FromDSHS\HS12CD.csv"),
  head = TRUE, sep = ",")

hosplink2 ← read.csv(file = str_c(dir, ":\users\scptex\Dropbox
  \Dissertation\Data\Hosp2012FromDSHS\HS12DIR.csv"),
  head = TRUE, sep = ",")

hosp ← merge(hosp, hosplink, by = "FID")

hosp ← merge(hosp, hosplink2, by = "FID")

source(str_c(dir, ":\users\scptex\Dropbox\Dissertation\Text
  \code\counties\LoadCountyFiles.r"))

counties$Name ← toupper(counties$Name)
counties$Name ← gsub(" COUNTY", "", counties$Name)
counties$Name ← gsub(" ", "", counties$Name)
hosp$county ← gsub(" ", "", hosp$COUNTY)

hosp_comerge ← merge(hosp, counties, by.x = "COUNTY", by.y = "
  Name")

hosp_comerge$MEDIANINCOME ← as.numeric(gsub(",", "", as.character
  (hosp_comerge$MEDIANINCOME)))

```

```

hosp_comerge$POVERTYPCT ← as.numeric(gsub(",", "", as.character(
  hosp_comerge$POVERTYPCT)))

HRRZipXref ← read.csv(file = str_c(dir, ":\users\\scptex\\
  Dropbox\\Dissertation\\Data\\\\DartmouthHRRs\\ZipHsaHrr11.csv"
),
  head = TRUE, sep = ",") #read in HRR-zip xref file

hosp_hrrmerge ← merge(hosp_comerge, HRRZipXref, by.x = "LOCOZIP",
  by.y = "zipcode11")

hosp_sub ← hosp_hrrmerge[, c("FID", "FacName", "EHR", "EHYear",
  "D1A1", "D1F1",
  "D1G1", "D2A11", "D2B11", "D2C11", "D2D11", "D3E1", "D5B1", "
  MSTAT", "POPDENSITY",
  "MEDIANINCOME", "POVERTYPCT", "hrrnum")]

names(hosp_sub) ← c("fid", "facname", "ehr", "ehryear", "
  totalbeds", "inpatientdays",
  "edvisits", "medicaredischarges", "medicaredays", "
  medicaiddischarges",
  "medicaiddays", "totalrevenue", "charitycare", "metrostat", "
  popdensity",
  "medianincome", "povertypct", "hrrnum")

hosp_sub$size ← cut(hosp_sub$totalbeds, labels = c("Small", "
  Medium", "Large"),
  breaks = c(0, 100, 400, 5000))

hosp_sub$metrostat ← gsub("1", "metro", hosp_sub$metrostat)
hosp_sub$metrostat ← gsub("2", "nonmetro", hosp_sub$metrostat)
hosp_sub$povpct_lvls ← cut(hosp_sub$povertypct, 4)
hosp_sub$metrostat ← factor(hosp_sub$metrostat)

hosp_sub$mcdpct ← hosp_sub$medicaiddays/hosp_sub$inpatientdays
hosp_sub$mcdehrelig ← ifelse(hosp_sub$mcdpct > 0.1, 1, 0)

hosp_sub1 ← hosp_sub[hosp_sub$ehr == 0, ]
hosp_sub2 ← hosp_sub[hosp_sub$ehr == 1 & hosp_sub$ehryear ≥
  2009, ]
hosp_sub ← rbind(hosp_sub1, hosp_sub2)

# *R Code Block* - mcdplotfittedcurve
hosp_sub$ehr ← as.numeric(hosp_sub$ehr) - 1

```

```
ggplot(hosp_sub[which(hosp_sub$mcdpct ≤ 1 & hosp_sub$mcdpct > 0),
], aes(x = mcdpct,
y = ehr)) + geom_point(position = position_jitter(width = 0.1,
height = 0.01),
alpha = 0.5) + geom_smooth(method = "loess") + geom_vline(
xintercept = 0.1,
linetype = "dotted") + xlim(0, 0.75)
```

```
# *R Code Block* - mcdbinscount
```

```
mcdhist ← hist(hosp_sub[which(hosp_sub$mcdpct < 1 & hosp_sub$mcdpct ≥ 0),
]$mcdpct, breaks = seq(0, 1, 0.01), plot = F)

mcd densfunc ← as.data.frame(cbind(mcdhist$mids, mcdhist$counts))
names(mcd densfunc) ← c("mids", "counts")

ggplot() + geom_histogram(data = hosp_sub[which(hosp_sub$mcdpct < 1),
], aes(x = mcdpct),
color = "black", fill = "white", binwidth = 0.01) + xlim(0, 1)
+ geom_vline(xintercept = 0.1,
linetype = "dotted") + geom_smooth(data = mcd densfunc, aes(x = mids, y = counts))
```

```
# *R Code Block* - calc mcd polyplots
```

```
hosp_sub$ehr ← as.numeric(hosp_sub$ehr)

# Imbens-Kalyanaraman Optimal Bandwidth Calculation
mcdIKbw ← IKbandwidth(hosp_sub$mcdpct, hosp_sub$ehr, cutpoint = 0.1)

# DCdensity(hosp_sub$mcdpct, cutpoint=0.1, bin=0.01)

mcd_lin_bw5 ← lm(ehr ~ mcdpct + mcdehrelig + mcdpct * mcdehrelig,
data = hosp_sub[which(hosp_sub$mcdpct < 0.15 & hosp_sub$mcdpct > 0.05),
])

mcd_lin_bw10 ← lm(ehr ~ mcdpct + mcdehrelig + mcdpct * mcdehrelig,
, data = hosp_sub[which(hosp_sub$mcdpct < 0.2),
])

mcd_quad ← lm(ehr ~ mcdpct + I(mcdpct^2) + mcdehrelig + mcdpct *
mcdehrelig +
I(mcdpct^2) * mcdehrelig, data = hosp_sub[which(hosp_sub$mcdpct < 0.4),
])
```

```

])

mcd_cub ← lm(ehr ~ mcdpct + I(mcdpct^2) + I(mcdpct^3) +
  mcdehrelig + mcdpct *
  mcdehrelig + I(mcdpct^2) * mcdehrelig + I(mcdpct^3) *
  mcdehrelig, data = hosp_sub[which(hosp_sub$mcdpct <
  0.4), ])

mcd_quart ← lm(ehr ~ mcdpct + I(mcdpct^2) + I(mcdpct^3) + I(
  mcdpct^4) + mcdehrelig +
  mcdpct * mcdehrelig + I(mcdpct^2) * mcdehrelig + I(mcdpct^3) *
  mcdehrelig +
  I(mcdpct^4) * mcdehrelig, data = hosp_sub[which(hosp_sub$
  mcdpct < 0.4),
  ])

mcd_disptbl_pt1 ← data.frame(x = seq(0, 0.1, by = 0.001), y1_5 =
  0, y1_10 = 0,
  y2 = 0, y3 = 0, y4 = 0)
mcd_disptbl_pt2 ← data.frame(x = seq(0.1, 0.2, by = 0.001), y1_5
  = 0, y1_10 = 0,
  y2 = 0, y3 = 0, y4 = 0)

coefs_mcd_lin_bw5 ← coefficients(mcd_lin_bw5)
coefs_mcd_lin_bw10 ← coefficients(mcd_lin_bw10)
coefs_mcd_quad ← coefficients(mcd_quad)
coefs_mcd_cub ← coefficients(mcd_cub)
coefs_mcd_quart ← coefficients(mcd_quart)

# calculate linear bw=5 estimates
mcd_disptbl_pt1$y1_5 ← coefs_mcd_lin_bw5[1] + coefs_mcd_lin_bw5
  [2] * mcd_disptbl_pt1$x
mcd_disptbl_pt2$y1_5 ← coefs_mcd_lin_bw5[1] + coefs_mcd_lin_bw5
  [2] * mcd_disptbl_pt2$x +
  coefs_mcd_lin_bw5[3]

# calculate linear bw=10 estimates
mcd_disptbl_pt1$y1_10 ← coefs_mcd_lin_bw10[1] +
  coefs_mcd_lin_bw10[2] * mcd_disptbl_pt1$x
mcd_disptbl_pt2$y1_10 ← coefs_mcd_lin_bw10[1] +
  coefs_mcd_lin_bw10[2] * mcd_disptbl_pt2$x +
  coefs_mcd_lin_bw10[3]

# calculate quadratic estimates
mcd_disptbl_pt1$y2 ← coefs_mcd_quad[1] + coefs_mcd_quad[2] *

```



```

med_disptbl_pt1$x +
  coefs_mcd_quad [3] * I(mcd_disptbl_pt1$x^2)
mcd_disptbl_pt2$y2 ← coefs_mcd_quad [1] + (coefs_mcd_quad [2] +
  coefs_mcd_quad [5]) *
  mcd_disptbl_pt2$x + (coefs_mcd_quad [3] + coefs_mcd_quad [6]) *
  I(mcd_disptbl_pt2$x^2) +
  coefs_mcd_quad [4]

# calculate cubic estimates
mcd_disptbl_pt1$y3 ← coefs_mcd_cub [1] + coefs_mcd_cub [2] *
  mcd_disptbl_pt1$x +
  coefs_mcd_cub [3] * I(mcd_disptbl_pt1$x^2) + coefs_mcd_cub [4] *
  I(mcd_disptbl_pt1$x^3)
mcd_disptbl_pt2$y3 ← coefs_mcd_cub [1] + (coefs_mcd_cub [2] +
  coefs_mcd_cub [6]) *
  mcd_disptbl_pt2$x + (coefs_mcd_cub [3] + coefs_mcd_cub [7]) * I(
  mcd_disptbl_pt2$x^2) +
  (coefs_mcd_cub [4] + coefs_mcd_cub [8]) * I(mcd_disptbl_pt2$x^3)
  + coefs_mcd_cub [5]

# calculate quartic estimates
mcd_disptbl_pt1$y4 ← coefs_mcd_quart [1] + coefs_mcd_quart [2] *
  mcd_disptbl_pt1$x +
  coefs_mcd_quart [3] * I(mcd_disptbl_pt1$x^2) + coefs_mcd_quart
  [4] * I(mcd_disptbl_pt1$x^3) +
  coefs_mcd_quart [5] * I(mcd_disptbl_pt1$x^4)
mcd_disptbl_pt2$y4 ← coefs_mcd_quart [1] + (coefs_mcd_quart [2] +
  coefs_mcd_quart [7]) *
  mcd_disptbl_pt2$x + (coefs_mcd_quart [3] + coefs_mcd_quart [8])
  * I(mcd_disptbl_pt2$x^2) +
  (coefs_mcd_quart [4] + coefs_mcd_quart [9]) * I(mcd_disptbl_pt2$
  x^3) + (coefs_mcd_quart [5] +
  coefs_mcd_quart [10]) * I(mcd_disptbl_pt2$x^4) +
  coefs_mcd_quart [6]

# *R Code Block* - dispmcdpolyplot
ggplot() + geom_line(data = mcd_disptbl_pt1, aes(x = x, y = y1.5,
  color = "Linear, bw=0.05")) +
  geom_line(data = mcd_disptbl_pt2, aes(x = x, y = y1.5, color =
  "Linear, bw=0.05")) +
  geom_line(data = mcd_disptbl_pt1, aes(x = x, y = y1.10, color
  = "Linear, bw=0.10")) +
  geom_line(data = mcd_disptbl_pt2, aes(x = x, y = y1.10, color
  = "Linear, bw=0.10")) +
  geom_smooth(data = mcd_disptbl_pt1, aes(x = x, y = y2, color =

```

```

    "Quadratic"),
    method = "lm", formula = y ~ x + I(x^2)) + geom_smooth(
      data = mcd_disptbl_pt2 ,
aes(x = x, y = y2, color = "Quadratic"), method = "lm",
  formula = y ~ x +
    I(x^2)) + geom_smooth(data = mcd_disptbl_pt1 , aes(x = x, y
      = y3, color = "Cubic"),
method = "lm", formula = y ~ x + I(x^2) + I(x^3)) +
  geom_smooth(data = mcd_disptbl_pt2 ,
aes(x = x, y = y3, color = "Cubic"), method = "lm", formula =
  y ~ x + I(x^2) +
    I(x^3)) + geom_smooth(data = mcd_disptbl_pt1 , aes(x = x, y
      = y4, color = "Quartic"),
method = "lm", formula = y ~ x + I(x^2) + I(x^3) + I(x^4)) +
  geom_smooth(data = mcd_disptbl_pt2 ,
aes(x = x, y = y4, color = "Quartic"), method = "lm", formula
  = y ~ x +
    I(x^2) + I(x^3) + I(x^4)) + geom_vline(xintercept = 0.1 ,
  linetype = "dotted") +
scale_color_discrete(name = "Model", breaks = c("Linear", bw=0
  .05", "Linear", bw=0.10",
  "Quadratic", "Cubic", "Quartic")) + geom_rect(mapping =
  aes(xmin = 0.075 ,
xmax = 0.125, ymin = 0, ymax = 1), fill = NA, color = "black",
  linetype = "dotted")

```

*# \*R Code Block\* - disp mcdpolyplotzoom*

```

ggplot() + geom_line(data = mcd_disptbl_pt1 , aes(x = x, y = y1_5 ,
  color = "Linear", bw=0.05")) +
  geom_line(data = mcd_disptbl_pt2 , aes(x = x, y = y1_5 , color =
    "Linear", bw=0.05")) +
  geom_line(data = mcd_disptbl_pt1 , aes(x = x, y = y1_10 , color
    = "Linear", bw=0.10")) +
  geom_line(data = mcd_disptbl_pt2 , aes(x = x, y = y1_10 , color
    = "Linear", bw=0.10")) +
  geom_smooth(data = mcd_disptbl_pt1 , aes(x = x, y = y2, color =
    "Quadratic"),
    method = "lm", formula = y ~ x + I(x^2)) + geom_smooth(
      data = mcd_disptbl_pt2 ,
aes(x = x, y = y2, color = "Quadratic"), method = "lm",
  formula = y ~ x +
    I(x^2)) + geom_smooth(data = mcd_disptbl_pt1 , aes(x = x, y
      = y3, color = "Cubic"),
method = "lm", formula = y ~ x + I(x^2) + I(x^3)) +
  geom_smooth(data = mcd_disptbl_pt2 ,

```

```

aes(x = x, y = y3, color = "Cubic"), method = "lm", formula =
  y ~ x + I(x^2) +
  I(x^3)) + geom_smooth(data = mcd_disptbl_pt1, aes(x = x, y
    = y4, color = "Quartic"),
  method = "lm", formula = y ~ x + I(x^2) + I(x^3) + I(x^4)) +
  geom_smooth(data = mcd_disptbl_pt2,
  aes(x = x, y = y4, color = "Quartic"), method = "lm", formula
    = y ~ x +
    I(x^2) + I(x^3) + I(x^4)) + geom_vline(xintercept = 0.1,
    linetype = "dotted") +
  scale_color_discrete(name = "Model", breaks = c("Linear", bw=0
    .05", "Linear", bw=0.10",
    "Quadratic", "Cubic", "Quartic")) + xlim(0.075, 0.125) +
  ylim(0, 1)

```

```

# *R Code Block* - mcdefftbl
mcd_eff_bw5 <- summary(mcd_lin_bw5)$coefficients[3, c(1, 4)]
mcd_eff_bw10 <- summary(mcd_lin_bw10)$coefficients[3, c(1, 4)]
mcd_eff_quad <- summary(mcd_quad)$coefficients[4, c(1, 4)]
mcd_eff_cub <- summary(mcd_cub)$coefficients[5, c(1, 4)]
mcd_eff_quart <- summary(mcd_quart)$coefficients[6, c(1, 4)]

mcd_eff <- rbind(mcd_eff_bw5, mcd_eff_bw10, mcd_eff_quad,
  mcd_eff_cub, mcd_eff_quart)
rownames(mcd_eff) <- c("Local linear, bandwidth=0.05", "Local
  linear, bandwidth=0.10",
  "Quadratic", "Cubic", "Quartic")

xtable(mcd_eff, type = "latex", align = "|r|r|r|", asis = "TRUE",
  floating = FALSE,
  digits = 4, label = "mcd_eff", caption = "Medicaid RDD Effect
  Sizes at Cut-off")

```

## K.5 R Code Used for Statistics, Tables, and Figures in Chapter 8 - Exploring the attitudes and motivations of health care providers regarding electronic health records and federal programs to promote their adoption

```
# *R Code Block* - loadlibraries - Load non-base R libraries
rm(list = ls())
library(stringr)
library(xtable)
library(ggplot2)
library(plyr)
library(corrgram)
library(stargazer)
library(BaylorEdPsych)
library(sas7bdat)
```

```
# *R Code Block* - init
figtheme ← theme(plot.title = element_text(face = "bold", size =
  16)) + theme(axis.title.y = element_text(face = "bold",
  size = 14), axis.text.y = element_text(size = 12, color = "
  black"), axis.title.x = element_text(face = "bold",
  size = 14), axis.text.x = element_text(size = 12, color = "
  black"))

dir ← "C"
```

```
# *R Code Block* - loadpractfile
source(str_c(dir, ":\users\\sctex\\Dropbox\\Dissertation\\Text\\
  code\\logit_practfile\\Pract2012DataLoad.r"))

pract$EHR ← pract$EHR1

pract$EHR ← sub(1, 1, pract$EHR)
pract$EHR ← sub(2, 1, pract$EHR)
pract$EHR ← sub(3, 0, pract$EHR)
pract$EHR ← sub(4, 0, pract$EHR)
pract$EHR ← sub(5, 0, pract$EHR)
```

```
# *R Code Block* - makephyslikdisptbl
tbl ← data.frame()

test ← pract[!(pract$EHR9_A %in% c(88, 99)), c("EHR", "EHR9_A")]
```

```

tmptbl ← as.data.frame(t(rbind(table(test$EHR9_A)/sum(table(test$
  EHR9_A)),
  names(table(test$EHR9_A))))
tmptbl$q ← "qA"
tbl ← rbind(tbl, tmptbl)

test ← pract[!(pract$EHR9_B %in% c(88, 99)), c("EHR", "EHR9_B")]
tmptbl ← as.data.frame(t(rbind(table(test$EHR9_B)/sum(table(test$
  EHR9_B)),
  names(table(test$EHR9_B))))
tmptbl$q ← "qB"
tbl ← rbind(tbl, tmptbl)

test ← pract[!(pract$EHR9_C %in% c(88, 99)), c("EHR", "EHR9_C")]
tmptbl ← as.data.frame(t(rbind(table(test$EHR9_C)/sum(table(test$
  EHR9_C)),
  names(table(test$EHR9_C))))
tmptbl$q ← "qC"
tbl ← rbind(tbl, tmptbl)

test ← pract[!(pract$EHR9_D %in% c(88, 99)), c("EHR", "EHR9_D")]
tmptbl ← as.data.frame(t(rbind(table(test$EHR9_D)/sum(table(test$
  EHR9_D)),
  names(table(test$EHR9_D))))
tmptbl$q ← "qD"
tbl ← rbind(tbl, tmptbl)

test ← pract[!(pract$EHR9_E %in% c(88, 99)), c("EHR", "EHR9_E")]
tmptbl ← as.data.frame(t(rbind(table(test$EHR9_E)/sum(table(test$
  EHR9_E)),
  names(table(test$EHR9_E))))
tmptbl$q ← "qE"
tbl ← rbind(tbl, tmptbl)

test ← pract[!(pract$EHR9_F %in% c(88, 99)), c("EHR", "EHR9_F")]
tmptbl ← as.data.frame(t(rbind(table(test$EHR9_F)/sum(table(test$
  EHR9_F)),
  names(table(test$EHR9_F))))
tmptbl$q ← "qF"
tbl ← rbind(tbl, tmptbl)

test ← pract[!(pract$EHR9_G %in% c(88, 99)), c("EHR", "EHR9_G")]
tmptbl ← as.data.frame(t(rbind(table(test$EHR9_G)/sum(table(test$
  EHR9_G)),
  names(table(test$EHR9_G))))

```

```

tmptbl$q ← "qG"
tbl ← rbind(tbl, tmptbl)

test ← pract[!(pract$EHR9_H %in% c(88, 99)), c("EHR", "EHR9_H")]
tmptbl ← as.data.frame(t(rbind(table(test$EHR9_H)/sum(table(test$
  EHR9_H)),
  names(table(test$EHR9_H))))))
tmptbl$q ← "qH"
tbl ← rbind(tbl, tmptbl)

test ← pract[!(pract$EHR9_I %in% c(88, 99)), c("EHR", "EHR9_I")]
tmptbl ← as.data.frame(t(rbind(table(test$EHR9_I)/sum(table(test$
  EHR9_I)),
  names(table(test$EHR9_I))))))
tmptbl$q ← "qI"
tbl ← rbind(tbl, tmptbl)

test ← pract[!(pract$EHR9_J %in% c(88, 99)), c("EHR", "EHR9_J")]
tmptbl ← as.data.frame(t(rbind(table(test$EHR9_J)/sum(table(test$
  EHR9_J)),
  names(table(test$EHR9_J))))))
tmptbl$q ← "qJ"
tbl ← rbind(tbl, tmptbl)

test ← pract[!(pract$EHR9_K %in% c(88, 99)), c("EHR", "EHR9_K")]
tmptbl ← as.data.frame(t(rbind(table(test$EHR9_K)/sum(table(test$
  EHR9_K)),
  names(table(test$EHR9_K))))))
tmptbl$q ← "qK"
tbl ← rbind(tbl, tmptbl)

test ← pract[!(pract$EHR9_L %in% c(88, 99)), c("EHR", "EHR9_L")]
tmptbl ← as.data.frame(t(rbind(table(test$EHR9_L)/sum(table(test$
  EHR9_L)),
  names(table(test$EHR9_L))))))
tmptbl$q ← "qL"
tbl ← rbind(tbl, tmptbl)

test ← pract[!(pract$EHR9_M %in% c(88, 99)), c("EHR", "EHR9_M")]
tmptbl ← as.data.frame(t(rbind(table(test$EHR9_M)/sum(table(test$
  EHR9_M)),
  names(table(test$EHR9_M))))))
tmptbl$q ← "qM"
tbl ← rbind(tbl, tmptbl)

```

```

test ← pract[!(pract$EHR9_N %in% c(88, 99)), c("EHR", "EHR9_N")]
tmptbl ← as.data.frame(t(rbind(table(test$EHR9_N)/sum(table(test$
  EHR9_N)),
  names(table(test$EHR9_N))))
tmptbl$q ← "qN"
tbl ← rbind(tbl, tmptbl)

tbl ← tbl[, c("V1", "V2", "q")]
names(tbl) ← c("Pct", "Response", "Question")
rownames(tbl) ← 1:nrow(tbl)

tbl$Pct ← as.character(tbl$Pct)
tbl$Pct ← as.numeric(tbl$Pct)
# tbl$Question ← factor(tbl$Question, levels=tbl[order(tbl$Question)
  ])

wraplength ← 25

txtA ← "The system facilitates communication of patient
  information among members of our health care team"
txtB ← "I feel the use of the system improves the quality of
  patient care"
txtC ← "I feel confident in my ability to assist others in using
  the system "
txtD ← "The system is more efficient than the old way of doing
  things"
txtE ← "I am physically comfortable with using the systems
  equipment "
txtF ← "I am satisfied with the mechanism for identifying/
  reporting issues that need to be fixed in the system."
txtG ← "With the system, patient information is more confidential
  /secure."
txtH ← "Information almost never gets lost in the system."
txtI ← "The use of the system reduces errors."
txtJ ← "Overall, I prefer using the system than the old way of
  doing things."
txtK ← "The system allows me to spend more time on other aspects
  of patient care."
txtL ← "The system takes into account the specific needs of my
  care area(s)."
txtM ← "When the system is unavailable, the backup way of doing
  things works adequately."
txtN ← "Information from the system enables me to make better
  decisions about patient care."

```

```

tmpA ← strwrap(txtA, wraplength, simplify = F)
tmpB ← strwrap(txtB, wraplength, simplify = F)
tmpC ← strwrap(txtC, wraplength, simplify = F)
tmpD ← strwrap(txtD, wraplength, simplify = F)
tmpE ← strwrap(txtE, wraplength, simplify = F)
tmpF ← strwrap(txtF, wraplength, simplify = F)
tmpG ← strwrap(txtG, wraplength, simplify = F)
tmpH ← strwrap(txtH, wraplength, simplify = F)
tmpI ← strwrap(txtI, wraplength, simplify = F)
tmpJ ← strwrap(txtJ, wraplength, simplify = F)
tmpK ← strwrap(txtK, wraplength, simplify = F)
tmpL ← strwrap(txtL, wraplength, simplify = F)
tmpM ← strwrap(txtM, wraplength, simplify = F)
tmpN ← strwrap(txtN, wraplength, simplify = F)

wrapA ← lapply(tmpA, paste, collapse = "\n")
wrapB ← lapply(tmpB, paste, collapse = "\n")
wrapC ← lapply(tmpC, paste, collapse = "\n")
wrapD ← lapply(tmpD, paste, collapse = "\n")
wrapE ← lapply(tmpE, paste, collapse = "\n")
wrapF ← lapply(tmpF, paste, collapse = "\n")
wrapG ← lapply(tmpG, paste, collapse = "\n")
wrapH ← lapply(tmpH, paste, collapse = "\n")
wrapI ← lapply(tmpI, paste, collapse = "\n")
wrapJ ← lapply(tmpJ, paste, collapse = "\n")
wrapK ← lapply(tmpK, paste, collapse = "\n")
wrapL ← lapply(tmpL, paste, collapse = "\n")
wrapM ← lapply(tmpM, paste, collapse = "\n")
wrapN ← lapply(tmpN, paste, collapse = "\n")

tbl$QuestionLabel ← tbl$Question

tbl$Question ← sub("qA", wrapA, tbl$Question)
tbl$Question ← sub("qB", wrapB, tbl$Question)
tbl$Question ← sub("qC", wrapC, tbl$Question)
tbl$Question ← sub("qD", wrapD, tbl$Question)
tbl$Question ← sub("qE", wrapE, tbl$Question)
tbl$Question ← sub("qF", wrapF, tbl$Question)
tbl$Question ← sub("qG", wrapG, tbl$Question)
tbl$Question ← sub("qH", wrapH, tbl$Question)
tbl$Question ← sub("qI", wrapI, tbl$Question)
tbl$Question ← sub("qJ", wrapJ, tbl$Question)
tbl$Question ← sub("qK", wrapK, tbl$Question)
tbl$Question ← sub("qL", wrapL, tbl$Question)
tbl$Question ← sub("qM", wrapM, tbl$Question)

```



```
tbl$Question ← sub("qN", wrapN, tbl$Question)

# *R Code Block* - makephysliksubtbl_quality
subtbl1 ← tbl[tbl$QuestionLabel %in% c("qA", "qB", "qI", "qN"), ]
# couldn't get the transform to work subtbl1 ← transform(subtbl1,
  Question
# = reorder(Question, Pct))

# *R Code Block* - plotphysliksubtbl_quality
ggplot(data = subtbl1, aes(x = Question, y = Pct, fill = factor(
  Response))) +
  geom_bar(width = 0.5, stat = "identity", color = "black") +
  coord_flip() +
  scale_fill_discrete(name = "Response", breaks = c("1", "2", "3",
    "4", "5",
    "6"), labels = c("Strongly Agree", "Agree", "Somewhat
    Agree", "Somewhat Disagree",
    "Disagree", "Strongly Disagree")) + figtheme

# *R Code Block* - makephysliksubtbl_usability
subtbl2 ← tbl[tbl$QuestionLabel %in% c("qC", "qD", "qE", "qJ"), ]

# *R Code Block* - plotphysliksubtbl_usability
ggplot(data = subtbl2, aes(x = Question, y = Pct, fill = factor(
  Response))) +
  geom_bar(width = 0.5, stat = "identity", color = "black") +
  coord_flip() +
  scale_fill_discrete(name = "Response", breaks = c("1", "2", "3",
    "4", "5",
    "6"), labels = c("Strongly Agree", "Agree", "Somewhat
    Agree", "Somewhat Disagree",
    "Disagree", "Strongly Disagree")) + figtheme

# *R Code Block* - makenonadopterreasonsdf
nos ← data.frame()
nos1 ← pract$EHR12.1[!(pract$EHR12.1 %in% c("777", "888", "999",
  NA, "11"))]
nos2 ← pract$EHR12.2[!(pract$EHR12.2 %in% c("777", "888", "999",
  NA, "11"))]
nos3 ← pract$EHR12.3[!(pract$EHR12.3 %in% c("777", "888", "999",
  NA, "11"))]
nos4 ← pract$EHR12.4[!(pract$EHR12.4 %in% c("777", "888", "999",
  NA, "11"))]
nos5 ← pract$EHR12.5[!(pract$EHR12.5 %in% c("777", "888", "999",
  NA, "11"))]
```

```

nos6 ← pract$EHR12_6[!(pract$EHR12_6 %in% c("777", "888", "999",
  NA, "11"))]
nos7 ← pract$EHR12_7[!(pract$EHR12_7 %in% c("777", "888", "999",
  NA, "11"))]
nos8 ← pract$EHR12_8[!(pract$EHR12_8 %in% c("777", "888", "999",
  NA, "11"))]
nos9 ← pract$EHR12_9[!(pract$EHR12_9 %in% c("777", "888", "999",
  NA, "11"))]
nos10 ← pract$EHR12_10[!(pract$EHR12_10 %in% c("777", "888", "999",
  ", NA, "11"))]

nos ← c(nos1, nos2, nos3, nos4, nos5, nos6, nos7, nos8, nos9,
  nos10)

nos_df ← as.data.frame(table(nos))

nos_df$nos ← str_c("n", nos_df$nos, "n")

wraplength2 ← 30

no1 ← "Software cost"
no2 ← "Hardware cost"
no3 ← "Personnel cost"
no4 ← "Return on investment"
no5 ← "Participation from physicians"
no6 ← "Organizational culture"
no7 ← "Interoperability with other electronic systems"
no8 ← "Lack of customization for specific needs"
no9 ← "Privacy/security concerns"
no10 ← "Legal barriers"

tmp1 ← strwrap(no1, wraplength2, simplify = F)
tmp2 ← strwrap(no2, wraplength2, simplify = F)
tmp3 ← strwrap(no3, wraplength2, simplify = F)
tmp4 ← strwrap(no4, wraplength2, simplify = F)
tmp5 ← strwrap(no5, wraplength2, simplify = F)
tmp6 ← strwrap(no6, wraplength2, simplify = F)
tmp7 ← strwrap(no7, wraplength2, simplify = F)
tmp8 ← strwrap(no8, wraplength2, simplify = F)
tmp9 ← strwrap(no9, wraplength2, simplify = F)
tmp10 ← strwrap(no10, wraplength2, simplify = F)

wrap1 ← lapply(tmp1, paste, collapse = "\n")
wrap2 ← lapply(tmp2, paste, collapse = "\n")
wrap3 ← lapply(tmp3, paste, collapse = "\n")

```

```

wrap4 ← lapply(tmp4, paste, collapse = "\n")
wrap5 ← lapply(tmp5, paste, collapse = "\n")
wrap6 ← lapply(tmp6, paste, collapse = "\n")
wrap7 ← lapply(tmp7, paste, collapse = "\n")
wrap8 ← lapply(tmp8, paste, collapse = "\n")
wrap9 ← lapply(tmp9, paste, collapse = "\n")
wrap10 ← lapply(tmp10, paste, collapse = "\n")

nos_df$noslabel ← nos_df$nos

nos_df$nos ← sub("n1n", wrap1, nos_df$nos)
nos_df$nos ← sub("n2n", wrap2, nos_df$nos)
nos_df$nos ← sub("n3n", wrap3, nos_df$nos)
nos_df$nos ← sub("n4n", wrap4, nos_df$nos)
nos_df$nos ← sub("n5n", wrap5, nos_df$nos)
nos_df$nos ← sub("n6n", wrap6, nos_df$nos)
nos_df$nos ← sub("n7n", wrap7, nos_df$nos)
nos_df$nos ← sub("n8n", wrap8, nos_df$nos)
nos_df$nos ← sub("n9n", wrap9, nos_df$nos)
nos_df$nos ← sub("n10n", wrap10, nos_df$nos)

nos_df$Freq ← nos_df$Freq/1477

# *R Code Block* - plotnonadopterreasonsfig
nos_df$nos ← factor(nos_df$nos)

nos_df ← transform(nos_df, nos = reorder(nos, Freq))

ggplot(data = nos_df, aes(x = nos, y = Freq, fill = nos)) +
  geom_bar(width = 0.5,
    stat = "identity", color = "black") + figtheme + coord_flip()
  + theme(legend.position = "none") +
  xlab("Reasons for Not Adopting\n") + ylab("\nPercent of
  Non-Adopters")

# *R Code Block* - muquestionfig
tmp ← data.frame(table(pract[!(pract$EHR %in% c(88, 99)) & !(
  pract$EHRI2 %in%
  c(88, 99))], c("EHR", "EHRI2"))))

tmp1 ← tmp[tmp$EHR == 0, ]
tmp2 ← tmp[tmp$EHR == 1, ]

tmp1$Freq ← tmp1$Freq/sum(tmp1$Freq)
tmp2$Freq ← tmp2$Freq/sum(tmp2$Freq)

```

```

tmp ← rbind(tmp1, tmp2)

tmp$EHRI2 ← sub(1, "Yes", tmp$EHRI2)
tmp$EHRI2 ← sub(2, "No", tmp$EHRI2)

tmp$EHR ← sub(0, "Non-Adopter", tmp$EHR)
tmp$EHR ← sub(1, "Adopter", tmp$EHR)

legendlabel ← "Did or will your practice seek incentive payments
to adopt, implement, and upgrade a certified electronic health
record system?"
lgndwrap ← 30
legendlabelwrap ← lapply(strwrap(legendlabel, lgndwrap, simplify
= F), paste,
collapse = "\n")

ggplot(data = tmp, aes(x = EHR, y = Freq, fill = EHRI2)) +
geom_bar(width = 0.5,
stat = "identity", color = "black") + figtheme +
scale_fill_discrete(name = legendlabelwrap) +
xlab("\nEHR Adoption Status") + ylab("Percent Seeking
Incentives\n")

```

```

# *R Code Block* - hosponadoptreasons
hosp ← read.csv(file = str_c(dir, ":\users\\scptex\\Dropbox\\
Dissertation\\Data\\2012HospSurvey\\HospSurv2012.csv"),
head = TRUE, sep = ",")

hosp$EHR_orig ← hosp$EHR
hosp$EHR ← sub(2, 1, hosp$EHR)
hosp$EHR ← sub(3, 0, hosp$EHR)
hosp$EHR ← sub(4, 0, hosp$EHR)
hosp$EHR ← sub(5, 0, hosp$EHR)
hosp$EHR ← sub(6, 0, hosp$EHR)
hosp$EHR ← as.factor(hosp$EHR)

hosp_nos ← hosp[!is.na(hosp$EHR), c("EHR", "EHR_orig", "Softcost",
"Hardcost",
"Perscost", "ROI", "Phypart", "Orgcult", "Interop", "Custom",
"Privsec",
"Legal", "Other")]

hosp_nos$Softcost[is.na(hosp_nos$Softcost)] ← 0
hosp_nos$Hardcost[is.na(hosp_nos$Hardcost)] ← 0

```

```

hosp_nos$Perscost[is.na(hosp_nos$Perscost)] ← 0
hosp_nos$ROI[is.na(hosp_nos$ROI)] ← 0
hosp_nos$Phypart[is.na(hosp_nos$Phypart)] ← 0
hosp_nos$Orgcult[is.na(hosp_nos$Orgcult)] ← 0
hosp_nos$Interop[is.na(hosp_nos$Interop)] ← 0
hosp_nos$Custom[is.na(hosp_nos$Custom)] ← 0
hosp_nos$Privsec[is.na(hosp_nos$Privsec)] ← 0
hosp_nos$Legal[is.na(hosp_nos$Legal)] ← 0

hnossums ← data.frame(apply(hosp_nos[, 3:12], 2, sum))

names(hnossums) ← "count"

barrier1 ← "Software Cost"
barrier2 ← "Hardware Cost"
barrier3 ← "Personnel Cost"
barrier4 ← "Return on Investment"
barrier5 ← "Physician Participation"
barrier6 ← "Organizational Culture"
barrier7 ← "Interoperability with Other Systems"
barrier8 ← "Lack of Customization for Specific Needs"
barrier9 ← "Privacy/Security Concerns"
barrier10 ← "Legal Barriers"

wrp ← 15

tmp1 ← strwrap(barrier1, wrp, simplify = F)
tmp2 ← strwrap(barrier2, wrp, simplify = F)
tmp3 ← strwrap(barrier3, wrp, simplify = F)
tmp4 ← strwrap(barrier4, wrp, simplify = F)
tmp5 ← strwrap(barrier5, wrp, simplify = F)
tmp6 ← strwrap(barrier6, wrp, simplify = F)
tmp7 ← strwrap(barrier7, wrp, simplify = F)
tmp8 ← strwrap(barrier8, wrp, simplify = F)
tmp9 ← strwrap(barrier9, wrp, simplify = F)
tmp10 ← strwrap(barrier10, wrp, simplify = F)

wrap1 ← lapply(tmp1, paste, collapse = "\n")
wrap2 ← lapply(tmp2, paste, collapse = "\n")
wrap3 ← lapply(tmp3, paste, collapse = "\n")
wrap4 ← lapply(tmp4, paste, collapse = "\n")
wrap5 ← lapply(tmp5, paste, collapse = "\n")
wrap6 ← lapply(tmp6, paste, collapse = "\n")
wrap7 ← lapply(tmp7, paste, collapse = "\n")
wrap8 ← lapply(tmp8, paste, collapse = "\n")

```

```

wrap9 ← lapply(tmp9, paste, collapse = "\n")
wrap10 ← lapply(tmp10, paste, collapse = "\n")

hnossums$barrier ← as.character(c(wrap1, wrap2, wrap3, wrap4,
  wrap5, wrap6,
  wrap7, wrap8, wrap9, wrap10))

axislab ← "Count of Hospitals with No Plans to Adopt EHRs (n=7)
  Reporting Each Barrier"
axiswrp ← 45
axislab2 ← strwrap(axislab, axiswrp, simplify = F)
axislab3 ← as.character(lapply(axislab2, paste, collapse = "\n"))

# *R Code Block* - hosponadoptplot
hnossums ← transform(hnossums, barrier = reorder(barrier, count))

ggplot(data = hnossums, aes(x = barrier, y = count, fill = barrier
)) + geom_bar(width = 0.5,
  stat = "identity", color = "black") + figtheme + coord_flip()
  + theme(legend.position = "none") +
  xlab("Barriers to EHR Implementation\n") + ylab(axislab3)

# *R Code Block* - hospmuquestionfig
tmp ← data.frame(table(hosp[!(hosp$Incent == 25), c("EHR", "
  Incent")]))

tmp1 ← tmp[tmp$EHR == 0, ]
tmp2 ← tmp[tmp$EHR == 1, ]

tmp1$Freq ← tmp1$Freq/sum(tmp1$Freq)
tmp2$Freq ← tmp2$Freq/sum(tmp2$Freq)

tmp ← rbind(tmp1, tmp2)

tmp$Incent ← sub(1, "Yes", tmp$Incent)
tmp$Incent ← sub(2, "No", tmp$Incent)

tmp$EHR ← sub(0, "Non-Adopter", tmp$EHR)
tmp$EHR ← sub(1, "Adopter", tmp$EHR)

legendlabel ← "Is facility seeking incentive payments for the
  adoption and meaningful use of EHRs?"
lgndwrap ← 30
legendlabelwrap ← lapply(strwrap(legendlabel, lgndwrap, simplify
  = F), paste,

```

```
collapse = "\n")

ggplot(data = tmp, aes(x = EHR, y = Freq, fill = Incent)) +
  geom_bar(width = 0.5,
    stat = "identity", color = "black") + figtheme +
    scale_fill_discrete(name = legendlabelwrap) +
  xlab("\nEHR Adoption Status") + ylab("Percent Seeking
  Incentives\n")
```

## K.6 R Code Used for Statistics, Tables, and Figures in Appendix A - Physician Logit Analysis Including Hospital Referral Regions

```
# This document has been 'knit' together to include R code, R  
output,  
# narrative text, figures, and tables.
```

```
# *R Code Block* - loadlibraries - Load non-base R libraries  
rm(list = ls())  
library(stringr)  
library(xtable)  
library(ggplot2)  
library(plyr)  
library(corrgram)  
library(stargazer)  
library(BaylorEdPsych)  
library(sas7bdat)
```

```
# *R Code Block* - setdefaults - Set R defaults and constants
```

```
# options(width=50)
```

```
figtheme ← theme(plot.title = element_text(face = "bold", size =  
16)) + theme(axis.title.y = element_text(face = "bold",  
size = 14), axis.text.y = element_text(size = 12, color = "  
black"), axis.title.x = element_text(face = "bold",  
size = 14), axis.text.x = element_text(size = 12, color = "  
black"))
```

```
options(bitmapType = "cairo")
```

```
# *R Code Block* - printtbl - Load my function to print survey  
question  
# answer options for a particular question
```

```
PrintTbl ← function(CodeBook, SurveyField) {  
  
  c2 ← CodeBook[CodeBook$Name == SurveyField, c("Question", "  
Label", "Value")]  
  c2tab ← c2[c2$Question == 0, c("Label", "Value")]  
  c2lab ← c2[c2$Question == 1, "Label"]  
  tabout ← xtable(c2tab, label = SurveyField, caption = c2lab,  
align = "|l|X|l|")  
}
```



```

print((tabout), include.rownames = FALSE, tabular.environment
      = "tabularx",
      width = "\\textwidth", type = "latex")
}

printxtab ← function(col1, col2, col1names, col2names, lbl, cap,
  aln) {
  testttab ← table(col1, col2)
  attr(testttab, "dimnames") ← list(col1names, col2names)
  xtable(testttab, type = "latex", asis = "TRUE", floating =
    FALSE, label = lbl,
    caption = cap, align = aln)
}

```

```

# *R Code Block* - loadpract11 - read pract 11 data and make pract
  2011 data
# frame
pract2011 ← read.csv(file = str_c(dir, ":\ users\scptex\dropbox
  \dissertation\data\2011practitionersurvey_work\
  practsurv2011.csv"),
  head = TRUE, sep = ",") #read in 2011 practitioners survey
  file
pract.sub2011 ← pract2011[pract2011$DI1 == 1 & pract2011$DI2 %in%
  c(1, 2) &
  !(pract2011$DI4 %in% c(8888, 9999)) & !(pract2011$EHR1 %in% c
  (88, 99)) &
  !(pract2011$DI3.COUNTY %in% c(88, 99)), c("resp_num", "DI2", "
  DI3.COUNTY.CODE1",
  "DI4", "EHR1", "HIE1", "EHR2.1", "EHR2.2", "EHR2.3", "EHR2.4",
  "EHR2.5",
  "EHR2.6", "EHR2.7", "EHR2.8", "EHR2.9", "EHR2.10", "EHR2.11",
  "EHR2.12",
  "EHR11", "EHR12", "EHR5", "DI3_ZIP")]

```

```

# *R Code Block* - loadpract12 - read pract 12 data and make pract
  2012 data
# frame
pract2012 ← read.csv(file = str_c(dir, ":\ users\scptex\dropbox
  \dissertation\data\2012\pract2012.csv"),
  head = TRUE, sep = ",")
pract.sub2012 ← pract2012[pract2012$DI1 == 1 & pract2012$DI2 %in%
  c(1, 2) &
  !(pract2012$DI4 %in% c(8888, 9999)) & !(pract2012$EHR1 %in% c
  (88, 99)) &
  !(pract2012$DI3.COUNTY %in% c(88, 99)), c("resp_num", "DI2", "

```

```

DI3_COUNTY_CODE_1_1",
"DI4", "EHR1", "HIE1", "EHR2_1", "EHR2_2", "EHR2_3", "EHR2_4",
"EHR2_5",
"EHR2_6", "EHR2_7", "EHR2_8", "EHR2_9", "EHR2_10", "EHR2_11",
"EHR2_12",
"EHR1", "EHR2", "EHR5", "DI3_ZIP") ]

```

```

# *R Code Block* - loadcountyfiles - codeblock to load countyxref
file and
# rename county field
counties <- read.csv(file = str_c(dir, ":\users\\scptex\\dropbox
\\dissertation\\data\\2011practitionersurvey_work\\
countyxref.csv"),
  head = TRUE, sep = ",") #read in counties file
# names(counties)[2]='COUNTYCODE' #rename county number field in
counties
counties$COUNTYCODE <- str_pad(as.character(counties$COUNTYCODE),
  3, side = "left",
  pad = "0")
counties$POPDENSITY <- as.numeric(gsub(",", "", as.character(
counties$Population)))/as.numeric(gsub(",",
"", as.character(counties$Size)))
countyincome <- read.csv(file = str_c(dir, ":\users\\scptex\\
dropbox\\dissertation\\data\\est10ALL.csv"),
  head = TRUE, sep = ",", skip = 2) #read in county income file
#load countyincome file, limit it to TX counties and
relevant columns, and format county FIPS field for
matching to other files
countyincome.tx <- countyincome[countyincome$Postal == "TX", c("
County.FIPS",
"Name", "Poverty.Estimate.All.Ages", "Poverty.Percent.All.Ages",
"Poverty.Estimate.Under.Age.18",
"Poverty.Percent.Under.Age.18", "Poverty.Estimate.Ages.5.17",
"Poverty.Percent.Ages.5.17",
"Median.Household.Income", "Poverty.Estimate.Ages.0.4", "
Poverty.Percent.Ages.0.4")]
countyincome.tx$County.FIPS <- str_pad(as.character(
countyincome.tx$County.FIPS),
  3, side = "left", pad = "0")
names(countyincome.tx)[which(names(countyincome.tx) == "
County.FIPS")] <- "COUNTYCODE"
names(countyincome.tx)[which(names(countyincome.tx) == "
Median.Household.Income")] <- "MEDIANINCOME"
names(countyincome.tx)[which(names(countyincome.tx) == "
Poverty.Percent.All.Ages")] <- "POVERTYPCT"

```

```

countyincome.tx$MEDIANINCOME ← as.numeric(gsub(","," ",
  as.character(countyincome.tx$MEDIANINCOME)))
countyincome.tx$POVERTYPCT ← as.numeric(gsub(","," ",
  as.character(countyincome.tx$POVERTYPCT)))
UrbanRural ← read.sas7bdat(str_c(dir, ":\ users\ scptex\ dropbox\
  \ dissertation\ data\ NCHSUrbanRural\
  NCHSUrbruralcodes.sas7bdat"))
UrbanRural ← UrbanRural[which(UrbanRural$ST_ABBREV == "TX"), ]
UrbanRural$CO_FIPS ← str_pad(as.character(UrbanRural$CO_FIPS), 3,
  side = "left",
  pad = "0")
urbrurxref ← as.data.frame(cbind(c(1, 2, 3, 4, 5, 6), c("Large
  central metro",
  "Large fringe metro", "Medium metro", "Small metro", "
  Micropolitan", "Noncore")))
names(urbrurxref) ← c("CODE2006", "MetroStatus")
counties ← merge(counties, countyincome.tx, by = "COUNTYCODE")
counties ← merge(counties, UrbanRural, by.x = "COUNTYCODE", by.y
  = "CO_FIPS")

```

```

# *R Code Block* - loadzipfile - load file with zipcode level
  poverty level
# stats (ACS)
zippov ← read.csv(file = str_c(dir, ":\ users\ scptex\ dropbox\
  dissertation\ Data\ ZIPPOV\ zippovpct.csv"),
  head = TRUE, sep = ",") #read in zipcode povpct xref
names(zippov)[which(names(zippov) == "povpct")] ← "ZIPPOVPCT"
zippov$ZIPPOVPCT ← as.numeric(as.character(zippov$ZIPPOVPCT))

```

```

# *R Code Block* - mergepractyears - merge 2011 and 2012
  practitioner files
pract.sub2011$survey_yr ← 0
pract.sub2012$survey_yr ← 1
names(pract.sub2011)[3] ← "COUNTYCODE"
names(pract.sub2012)[3] ← "COUNTYCODE"
pract.sub ← rbind(pract.sub2011, pract.sub2012)

```

```

# *R Code Block* - mergepractcounty - codeblock to merge
  practitioner and
# countyxref files, create new fields, re-code existing fields,
  create
# meaningful column names, and reduce data frame width for
  relevant columns
pract.sub$COUNTYCODE ← str_pad(as.character(pract.sub$COUNTYCODE)
  , 3, side = "left",

```

```

pad = "0")
# source(str_c(dir, ':\\users\\scptex\\Dropbox\\Dissertation\\Text
  \\code\\Counties\\LoadCountyFiles.r'))
pract.sub <- merge(pract.sub, counties, by = "COUNTYCODE") #merge
  pract.sub and counties on COUNTYCODE field
pract.sub$Population <- as.numeric(gsub(",", "", as.character(
  pract.sub$Population)))
pract.sub$Size <- as.numeric(gsub(",", "", as.character(pract.sub$
  Size))) #create new population density field
pract.sub$Population <- as.numeric(pract.sub$Population)
pract.sub$Size <- as.numeric(pract.sub$Size)
pract.sub$POPDENSITY <- pract.sub$Population/pract.sub$Size
pract.sub$DI2 <- as.integer(sub(2, 0, pract.sub$DI2)) #to sub
  zeroes for 2s in coding for DI2, Medicaid =1, no=0 after sub
names(pract.sub)[which(names(pract.sub) == "DI2")] <- "MEDICAID"
#rename DI2 to MEDICAID
names(pract.sub)[which(names(pract.sub) == "DI4")] <- "
  PRACTITIONERCOUNT" #rename DI4 to PRACTITIONERCOUNT
pract.sub$EHR1_orig <- pract.sub$EHR1
pract.sub$EHR1 <- sub(2, 1, pract.sub$EHR1) #code EHR1 2 to 1;
  makes paperless or mostly paperless 1
pract.sub$EHR1 <- sub(3, 0, pract.sub$EHR1) #code EHR1 3, 4, and
  5 to 0; make everything else 0
pract.sub$EHR1 <- sub(4, 0, pract.sub$EHR1)
pract.sub$EHR1 <- sub(5, 0, pract.sub$EHR1)
names(pract.sub)[which(names(pract.sub) == "EHR1")] <- "EHR" #
  rename EHR1 to EHR
pract.sub$EHR <- as.numeric(pract.sub$EHR) #make EHR a numeric
  field
pract.sub$EHR <- as.factor(pract.sub$EHR)
pract.sub$MEDICAID <- factor(pract.sub$MEDICAID) #make MEDICAID a
  factor; this didn't seem to change anything
pract.sub$CODE2006 <- factor(pract.sub$CODE2006)
pract.sub <- merge(pract.sub, urbrurxref, by = "CODE2006")
pract.sub$MetroStatus <- factor(pract.sub$MetroStatus, levels(
  pract.sub$MetroStatus)[c(5,
    4, 6, 3, 2, 1)])
pract.sub$popdens_lvls <- as.ordered(cut(pract.sub$POPDENSITY, 4))
pract.sub$practent_lvls <- cut(pract.sub$PRACTITIONERCOUNT, breaks
  = c(0, 1,
    5, 10, 100, 1000), labels = c("Solo (1)", "Small (2-5)", "
    Medium (6-10)",
    "Large (11-100)", "Very Large (101-1000)"))
pract.sub$medinc_lvls <- cut(pract.sub$MEDIANINCOME, 4)
pract.sub$povpct_lvls <- cut(pract.sub$POVERTYPCT, breaks = c(0,

```

```

10, 20, 30,
40), labels = c("Less than 10", "Between 10 and 20", "Between
20 and 30",
"Between 30 and 40"))
HRRZipXref ← read.csv(file = str_c(dir, ":\users\scptex\
Dropbox\Dissertation\Data\\\DartmouthHRRs\ZipHsaHrr11.csv"
),
head = TRUE, sep = ",") #read in HRR-zip xref file
pract.sub ← merge(pract.sub, HRRZipXref, by.x = "DI3-ZIP", by.y =
"zipcode11")

pract.sub ← merge(pract.sub, zippov, by.x = "DI3-ZIP", by.y = "
zip")
pract.sub$zippovpct_lvls ← cut(pract.sub$ZIPPOVPCT, breaks = c(0,
10, 20, 30,
40, 50, 60, 70), labels = c("Less than 10", "Between 10 and 20
", "Between 20 and 30",
"Between 30 and 40", "Between 40 and 50", "Between 50 and 60",
"Between 60 and 70"))

```

```

pract.sub$hrrnum ← factor(pract.sub$hrrnum)

firsttrymergedfacthrr ← glm(EHR ~ MEDICAID + practcnt_lvls +
survey_yr + hrrnum +
MEDICAID * survey_yr + practcnt_lvls * survey_yr + hrrnum *
survey_yr, data = pract.sub,
family = "binomial")
secondtrymergedfacthrr ← glm(EHR ~ MEDICAID + practcnt_lvls +
zippovpct_lvls +
survey_yr + hrrnum + MEDICAID * survey_yr + practcnt_lvls *
survey_yr +
zippovpct_lvls * survey_yr + hrrnum * survey_yr, data =
pract.sub, family = "binomial")
thirdtrymergedfacthrr ← glm(EHR ~ MEDICAID + practcnt_lvls +
zippovpct_lvls +
MetroStatus + survey_yr + hrrnum + MEDICAID * survey_yr +
practcnt_lvls *
survey_yr + zippovpct_lvls * survey_yr + MetroStatus *
survey_yr + hrrnum *
survey_yr, data = pract.sub, family = "binomial")

```

```

stargazer(firsttrymergedfacthrr, secondtrymergedfacthrr,
thirdtrymergedfacthrr,
title = "Physicians, Both Years, Factor Models with HRRs,
Table 1", no.space = TRUE,

```

```

    digits = 2, label = "mergedpractlogitfacthrr1", align = TRUE,
      keep = c(1:20),
      report = "vc*")

# , covariate.labels=c('Medicaid Provider','Practice Size - Small
# ', 'Practice
# Size - Medium','Practice Size - Large','Practice Size - Very
# Large',
# 'Year', 'HRR 219', 'HRR 382', 'HRR 383', 'HRR 385','HRR 386', '
# HRR
# 388', 'HRR 390', 'HRR 391', 'HRR 393', 'HRR 394', 'HRR 396', 'HRR
# 397', 'HRR
# 399', 'HRR 400')

stargazer(firsttrymergedfacthrr , secondtrymergedfacthrr ,
  thirdtrymergedfacthrr ,
  title = "Physicians, Both Years, Factor Models with HRRs,
  Table 2", no.space = TRUE,
  digits = 2, label = "mergedpractlogitfacthrr2", align = TRUE,
  keep = c(21:40),
  report = "vc*")

# , covariate.labels=c('HRR 402', 'HRR 406', 'HRR 411', 'HRR 412', '
# HRR
# 413', 'HRR 416', 'HRR 417', 'HRR 418', 'HRR 420', 'Poverty Pct
# (10-20)', 'Poverty Pct (20-30)', 'Poverty Pct (30-40)', 'Poverty
# Pct
# (40-50)', 'Poverty Pct (50-60)', 'Poverty Pct (60-70)', '
# Metro-Micro',
# 'Metro-Small', 'Metro-Medium', 'Metro-Large Fringe', '
# Metro-Large
# Central')

stargazer(firsttrymergedfacthrr , secondtrymergedfacthrr ,
  thirdtrymergedfacthrr ,
  title = "Physicians, Both Years, Factor Models with HRRs,
  Table 3", no.space = TRUE,
  digits = 2, label = "mergedpractlogitfacthrr3", align = TRUE,
  keep = c(41:60),
  report = "vc*")

stargazer(firsttrymergedfacthrr , secondtrymergedfacthrr ,
  thirdtrymergedfacthrr ,
  title = "Physicians, Both Years, Factor Models with HRRs,
  Table 4", no.space = TRUE,

```

```

digits = 2, label = "mergedpractlogitfacthrr4", align = TRUE,
  keep = c(61:80),
  report = "vc*")

# for displaying the city corresponding to HRR 406 --
# pract.sub[which(pract.sub$hrrnum=='406'),'hrrcity']

# , covariate.labels=c('Medicaid Provider','Practice Size - Small
# ', 'Practice
# Size - Medium','Practice Size - Large','Practice Size - Very
# Large','Poverty Pct (10-20)','Poverty Pct (20-30)','Poverty Pct
# (30-40)','Poverty Pct (40-50)','Poverty Pct (50-60)','Poverty
# Pct
# (60-70)', 'Metro-Micro', 'Metro-Small', 'Metro-Medium', '
# Metro-Large
# Fringe', 'Metro-Large Central')

```

## K.7 R Code Used for Statistics, Tables, and Figures in Appendix B - Physician EHR Adoption Estimated Using Linear Probability Models

```
# This document has been 'knit' together to include R code, R  
output,  
# narrative text, figures, and tables.
```

```
# *R Code Block* - loadlibraries - Load non-base R libraries  
rm(list = ls())  
library(stringr)  
library(xtable)  
library(ggplot2)  
library(plyr)  
library(corrgram)  
library(stargazer)  
library(BaylorEdPsych)  
library(sas7bdat)
```

```
# *R Code Block* - setdefaults - Set R defaults and constants
```

```
# options(width=50)
```

```
figtheme ← theme(plot.title = element_text(face = "bold", size =  
16)) + theme(axis.title.y = element_text(face = "bold",  
size = 14), axis.text.y = element_text(size = 12, color = "  
black"), axis.title.x = element_text(face = "bold",  
size = 14), axis.text.x = element_text(size = 12, color = "  
black"))
```

```
options(bitmapType = "cairo")
```

```
# *R Code Block* - printtbl - Load my function to print survey  
question
```

```
# answer options for a particular question
```

```
PrintTbl ← function(CodeBook, SurveyField) {  
  
  c2 ← CodeBook[CodeBook$Name == SurveyField, c("Question", "  
Label", "Value")]  
  c2tab ← c2[c2$Question == 0, c("Label", "Value")]  
  c2lab ← c2[c2$Question == 1, "Label"]  
  tabout ← xtable(c2tab, label = SurveyField, caption = c2lab,  
align = "|l|X|l|")  
}
```



```

print((tabout), include.rownames = FALSE, tabular.environment
      = "tabularx",
      width = "\\textwidth", type = "latex")
}

printxtab ← function(col1, col2, col1names, col2names, lbl, cap,
  aln) {
  testttab ← table(col1, col2)
  attr(testttab, "dimnames") ← list(col1names, col2names)
  xtable(testttab, type = "latex", asis = "TRUE", floating =
    FALSE, label = lbl,
    caption = cap, align = aln)
}

```

```

# *R Code Block* - loadpract11 - read pract 11 data and make pract
  2011 data
# frame
pract2011 ← read.csv(file = str_c(dir, ":\ users\scptex\dropbox
  \dissertation\data\2011practitionersurvey_work\
  practsurv2011.csv"),
  head = TRUE, sep = ",") #read in 2011 practitioners survey
  file
pract.sub2011 ← pract2011[pract2011$DI1 == 1 & pract2011$DI2 %in%
  c(1, 2) &
  !(pract2011$DI4 %in% c(8888, 9999)) & !(pract2011$EHR1 %in% c
  (88, 99)) &
  !(pract2011$DI3.COUNTY %in% c(88, 99)), c("resp_num", "DI2", "
  DI3.COUNTY.CODE1",
  "DI4", "EHR1", "HIE1", "EHR2.1", "EHR2.2", "EHR2.3", "EHR2.4",
  "EHR2.5",
  "EHR2.6", "EHR2.7", "EHR2.8", "EHR2.9", "EHR2.10", "EHR2.11",
  "EHR2.12",
  "EHR11", "EHR12", "EHR5", "DI3_ZIP")]

```

```

# *R Code Block* - loadpract12 - read pract 12 data and make pract
  2012 data
# frame
pract2012 ← read.csv(file = str_c(dir, ":\ users\scptex\dropbox
  \dissertation\data\2012\pract2012.csv"),
  head = TRUE, sep = ",")
pract.sub2012 ← pract2012[pract2012$DI1 == 1 & pract2012$DI2 %in%
  c(1, 2) &
  !(pract2012$DI4 %in% c(8888, 9999)) & !(pract2012$EHR1 %in% c
  (88, 99)) &
  !(pract2012$DI3.COUNTY %in% c(88, 99)), c("resp_num", "DI2", "

```

```

DI3_COUNTY_CODE_1_1",
"DI4", "EHR1", "HIE1", "EHR2_1", "EHR2_2", "EHR2_3", "EHR2_4",
"EHR2_5",
"EHR2_6", "EHR2_7", "EHR2_8", "EHR2_9", "EHR2_10", "EHR2_11",
"EHR2_12",
"EHR1", "EHR2", "EHR5", "DI3_ZIP") ]

```

```

# *R Code Block* - loadcountyfiles - codeblock to load countyxref
# file and
# rename county field
counties <- read.csv(file = str_c(dir, ":\users\\scptex\\dropbox
\\dissertation\\data\\2011practitionersurvey_work\\
countyxref.csv"),
  head = TRUE, sep = ",") #read in counties file
# names(counties)[2]='COUNTYCODE' #rename county number field in
# counties
counties$COUNTYCODE <- str_pad(as.character(counties$COUNTYCODE),
  3, side = "left",
  pad = "0")
counties$POPDENSITY <- as.numeric(gsub(",", "", as.character(
  counties$Population)))/as.numeric(gsub(",", "",
  "", as.character(counties$Size)))
countyincome <- read.csv(file = str_c(dir, ":\users\\scptex\\
dropbox\\dissertation\\data\\est10ALL.csv"),
  head = TRUE, sep = ",", skip = 2) #read in county income file
#load countyincome file, limit it to TX counties and
#relevant columns, and format county FIPS field for
#matching to other files
countyincome.tx <- countyincome[countyincome$Postal == "TX", c("
County.FIPS",
  "Name", "Poverty.Estimate.All.Ages", "Poverty.Percent.All.Ages",
  "Poverty.Estimate.Under.Age.18",
  "Poverty.Percent.Under.Age.18", "Poverty.Estimate.Ages.5.17",
  "Poverty.Percent.Ages.5.17",
  "Median.Household.Income", "Poverty.Estimate.Ages.0.4", "
  Poverty.Percent.Ages.0.4")]
countyincome.tx$County.FIPS <- str_pad(as.character(
  countyincome.tx$County.FIPS),
  3, side = "left", pad = "0")
names(countyincome.tx)[which(names(countyincome.tx) == "
County.FIPS")] <- "COUNTYCODE"
names(countyincome.tx)[which(names(countyincome.tx) == "
Median.Household.Income")] <- "MEDIANINCOME"
names(countyincome.tx)[which(names(countyincome.tx) == "
Poverty.Percent.All.Ages")] <- "POVERTYPCT"

```

```

countyincome.tx$MEDIANINCOME ← as.numeric(gsub(","," ",
  as.character(countyincome.tx$MEDIANINCOME)))
countyincome.tx$POVERTYPCT ← as.numeric(gsub(","," ",
  as.character(countyincome.tx$POVERTYPCT)))
UrbanRural ← read.sas7bdat(str_c(dir, ":\ users\ scptex\ dropbox
  \ dissertation \ data \ NCHSUrbanRural\
  NCHSUrbruralcodes.sas7bdat"))
UrbanRural ← UrbanRural[which(UrbanRural$ST_ABBREV == "TX"), ]
UrbanRural$CO_FIPS ← str_pad(as.character(UrbanRural$CO_FIPS), 3,
  side = "left",
  pad = "0")
urbrurxref ← as.data.frame(cbind(c(1, 2, 3, 4, 5, 6), c("Large
  central metro",
  "Large fringe metro", "Medium metro", "Small metro", "
  Micropolitan", "Noncore")))
names(urbrurxref) ← c("CODE2006", "MetroStatus")
counties ← merge(counties, countyincome.tx, by = "COUNTYCODE")
counties ← merge(counties, UrbanRural, by.x = "COUNTYCODE", by.y
  = "CO_FIPS")

```

```

# *R Code Block* - loadzipfile - load file with zipcode level
# poverty level
# stats (ACS)
zippov ← read.csv(file = str_c(dir, ":\ users\ scptex\ dropbox\
  dissertation \ Data \ ZIPPOV \ zippovpct.csv"),
  head = TRUE, sep = ",") #read in zipcode povpct xref
names(zippov)[which(names(zippov) == "povpct")] ← "ZIPPOVPCT"
zippov$ZIPPOVPCT ← as.numeric(as.character(zippov$ZIPPOVPCT))

```

```

# *R Code Block* - mergepractyears - merge 2011 and 2012
# practitioner files
pract.sub2011$survey_yr ← 0
pract.sub2012$survey_yr ← 1
names(pract.sub2011)[3] ← "COUNTYCODE"
names(pract.sub2012)[3] ← "COUNTYCODE"
pract.sub ← rbind(pract.sub2011, pract.sub2012)

```

```

# *R Code Block* - mergepractcounty - codeblock to merge
# practitioner and
# countyxref files, create new fields, re-code existing fields,
# create
# meaningful column names, and reduce data frame width for
# relevant columns
pract.sub$COUNTYCODE ← str_pad(as.character(pract.sub$COUNTYCODE)
  , 3, side = "left",

```

```

pad = "0")
# source(str_c(dir, ':\\users\\scptex\\Dropbox\\Dissertation\\Text
  \\code\\Counties\\LoadCountyFiles.r'))
pract.sub ← merge(pract.sub, counties, by = "COUNTYCODE") #merge
  pract.sub and counties on COUNTYCODE field
pract.sub$Population ← as.numeric(gsub(",", "", as.character(
  pract.sub$Population)))
pract.sub$Size ← as.numeric(gsub(",", "", as.character(pract.sub$
  Size))) #create new population density field
pract.sub$Population ← as.numeric(pract.sub$Population)
pract.sub$Size ← as.numeric(pract.sub$Size)
pract.sub$POPDENSITY ← pract.sub$Population/pract.sub$Size
pract.sub$DI2 ← as.integer(sub(2, 0, pract.sub$DI2)) #to sub
  zeroes for 2s in coding for DI2, Medicaid =1, no=0 after sub
names(pract.sub)[which(names(pract.sub) == "DI2")] ← "MEDICAID"
#rename DI2 to MEDICAID
names(pract.sub)[which(names(pract.sub) == "DI4")] ← "
  PRACTITIONERCOUNT" #rename DI4 to PRACTITIONERCOUNT
pract.sub$EHR1_orig ← pract.sub$EHR1
pract.sub$EHR1 ← sub(2, 1, pract.sub$EHR1) #code EHR1 2 to 1;
  makes paperless or mostly paperless 1
pract.sub$EHR1 ← sub(3, 0, pract.sub$EHR1) #code EHR1 3, 4, and
  5 to 0; make everything else 0
pract.sub$EHR1 ← sub(4, 0, pract.sub$EHR1)
pract.sub$EHR1 ← sub(5, 0, pract.sub$EHR1)
names(pract.sub)[which(names(pract.sub) == "EHR1")] ← "EHR" #
  rename EHR1 to EHR
pract.sub$EHR ← as.numeric(pract.sub$EHR) #make EHR a numeric
  field
pract.sub$EHR ← as.factor(pract.sub$EHR)
pract.sub$MEDICAID ← factor(pract.sub$MEDICAID) #make MEDICAID a
  factor; this didn't seem to change anything
pract.sub$CODE2006 ← factor(pract.sub$CODE2006)
pract.sub ← merge(pract.sub, urbrurxref, by = "CODE2006")
pract.sub$MetroStatus ← factor(pract.sub$MetroStatus, levels(
  pract.sub$MetroStatus)[c(5,
  4, 6, 3, 2, 1)])
pract.sub$popdens_lvls ← as.ordered(cut(pract.sub$POPDENSITY, 4))
pract.sub$practent_lvls ← cut(pract.sub$PRACTITIONERCOUNT, breaks
  = c(0, 1,
  5, 10, 100, 1000), labels = c("Solo (1)", "Small (2-5)", "
  Medium (6-10)",
  "Large (11-100)", "Very Large (101-1000)"))
pract.sub$medinc_lvls ← cut(pract.sub$MEDIANINCOME, 4)
pract.sub$povpct_lvls ← cut(pract.sub$POVERTYPCT, breaks = c(0,

```

```

10, 20, 30,
40), labels = c("Less than 10", "Between 10 and 20", "Between
20 and 30",
"Between 30 and 40"))
HRRZipXref ← read.csv(file = str_c(dir, ":\users\scptex\
Dropbox\Dissertation\Data\\\DartmouthHRRs\ZipHsaHrr11.csv"
),
head = TRUE, sep = ",") #read in HRR-zip xref file
pract.sub ← merge(pract.sub, HRRZipXref, by.x = "DI3-ZIP", by.y =
"zipcode11")

pract.sub ← merge(pract.sub, zippov, by.x = "DI3-ZIP", by.y = "
zip")
pract.sub$zippovpct_lvls ← cut(pract.sub$ZIPPOVPCT, breaks = c(0,
10, 20, 30,
40, 50, 60, 70), labels = c("Less than 10", "Between 10 and 20
", "Between 20 and 30",
"Between 30 and 40", "Between 40 and 50", "Between 50 and 60",
"Between 60 and 70"))

pract.sub$EHR ← as.numeric(pract.sub$EHR) - 1

firsttrymergedfactlpm ← glm(EHR ~ MEDICAID + practent_lvls +
survey_yr + MEDICAID *
survey_yr + practent_lvls * survey_yr, data = pract.sub,
family = "gaussian")
secondtrymergedfactlpm ← glm(EHR ~ MEDICAID + practent_lvls +
zippovpct_lvls +
survey_yr + MEDICAID * survey_yr + practent_lvls * survey_yr +
zippovpct_lvls *
survey_yr, data = pract.sub, family = "gaussian")
thirdtrymergedfactlpm ← glm(EHR ~ MEDICAID + practent_lvls +
zippovpct_lvls +
MetroStatus + survey_yr + MEDICAID * survey_yr + practent_lvls
* survey_yr +
zippovpct_lvls * survey_yr + MetroStatus * survey_yr, data =
pract.sub,
family = "gaussian")

stargazer(firsttrymergedfactlpm, secondtrymergedfactlpm,
thirdtrymergedfactlpm,
title = "Physicians, Both Years, Linear Probability Factor
Models, Explanatory Variables",
no.space = TRUE, digits = 2, label = "mergedpractlogitfact1",
align = TRUE,

```

```

keep = c(1:16), report = "vc*", covariate.labels = c("Medicaid
  Provider",
    "Practice Size - Small", "Practice Size - Medium", "
      Practice Size - Large",
    "Practice Size - Very Large", "Poverty Pct (10-20)", "
      Poverty Pct (20-30)",
    "Poverty Pct (30-40)", "Poverty Pct (40-50)", "Poverty Pct
      (50-60)",
    "Poverty Pct (60-70)", "Metro-Micro", "Metro-Small", "
      Metro-Medium",
    "Metro-Large Fringe", "Metro-Large Central"))

stargazer(firsttrymergedfactlpm, secondtrymergedfactlpm,
  thirdtrymergedfactlpm,
  title = "Physicians, Both Years, Linear Probability Factor
    Models (cont.), Year Effects",
  no.space = TRUE, digits = 2, label = "mergedpractlogitfact2",
  align = TRUE,
  keep = c(17:33), report = "vc*", covariate.labels = c("Survey
    Year", "Medicaid Provider*Year",
    "Practice Size - Small*Year", "Practice Size - Medium*Year
      ", "Practice Size - Large*Year",
    "Practice Size - Very Large*Year", "Poverty Pct (10-20)*
      Year", "Poverty Pct (20-30)*Year",
    "Poverty Pct (30-40)*Year", "Poverty Pct (40-50)*Year", "
      Poverty Pct (50-60)*Year",
    "Poverty Pct (60-70)*Year", "Metro-Micro*Year", "
      Metro-Small*Year",
    "Metro-Medium*Year", "Metro-Large Fringe*Year", "
      Metro-Large Central*Year",
    "Constant"))

```

## K.8 R Code Used for Statistics, Tables, and Figures in Appendix C - Hospital Logit Analysis Including Hospital Referral Regions

```
# This document has been 'knit' together to include R code, R  
output,  
# narrative text, figures, and tables.
```

```
# *R Code Block* - loadlibraries - Load non-base R libraries  
rm(list = ls())  
library(stringr)  
library(xtable)  
library(ggplot2)  
library(plyr)  
library(corrgram)  
library(stargazer)  
library(BaylorEdPsych)  
library(sas7bdat)
```

```
# *R Code Block* - setdefaults - Set R defaults and constants
```

```
# options(width=50)
```

```
figtheme ← theme(plot.title = element_text(face = "bold", size =  
16)) + theme(axis.title.y = element_text(face = "bold",  
size = 14), axis.text.y = element_text(size = 12, color = "  
black"), axis.title.x = element_text(face = "bold",  
size = 14), axis.text.x = element_text(size = 12, color = "  
black"))
```

```
options(bitmapType = "cairo")
```

```
# *R Code Block* - printtbl - Load my function to print survey  
question  
# answer options for a particular question
```

```
PrintTbl ← function(CodeBook, SurveyField) {  
  
  c2 ← CodeBook[CodeBook$Name == SurveyField, c("Question", "  
Label", "Value")]  
  c2tab ← c2[c2$Question == 0, c("Label", "Value")]  
  c2lab ← c2[c2$Question == 1, "Label"]  
  tabout ← xtable(c2tab, label = SurveyField, caption = c2lab,  
align = "|1|X|1|")  
}
```

```

    print((tabout), include.rownames = FALSE, tabular.environment
          = "tabularx",
          width = "\\textwidth", type = "latex")
}

printxtab ← function(col1, col2, col1names, col2names, lbl, cap,
  aln) {
  testttab ← table(col1, col2)
  attr(testttab, "dimnames") ← list(col1names, col2names)
  xtable(testttab, type = "latex", asis = "TRUE", floating =
    FALSE, label = lbl,
    caption = cap, align = aln)
}

# *R Code Block* - loadcountyfiles - codeblock to load countyxref
# file and
# rename county field
counties ← read.csv(file = str_c(dir, ":\users\\scptex\\dropbox
  \\dissertation\\data\\2011practitionersurvey_work\\
  countyxref.csv"),
  head = TRUE, sep = ",") #read in counties file
# names(counties)[2]='COUNTYCODE' #rename county number field in
# counties
counties$COUNTYCODE ← str_pad(as.character(counties$COUNTYCODE),
  3, side = "left",
  pad = "0")
counties$POPDENSITY ← as.numeric(gsub(",", "", as.character(
  counties$Population)))/as.numeric(gsub(",", "",
  "", as.character(counties$Size)))
countyincome ← read.csv(file = str_c(dir, ":\users\\scptex\\
  dropbox\\dissertation\\data\\est10ALL.csv"),
  head = TRUE, sep = ",", skip = 2) #read in county income file
#load countyincome file, limit it to TX counties and
#relevant columns, and format county FIPS field for
#matching to other files
countyincome.tx ← countyincome[countyincome$Postal == "TX", c("
  County.FIPS",
  "Name", "Poverty.Estimate.All.Ages", "Poverty.Percent.All.Ages",
  "Poverty.Estimate.Under.Age.18",
  "Poverty.Percent.Under.Age.18", "Poverty.Estimate.Ages.5.17",
  "Poverty.Percent.Ages.5.17",
  "Median.Household.Income", "Poverty.Estimate.Ages.0.4", "
  Poverty.Percent.Ages.0.4")]
countyincome.tx$County.FIPS ← str_pad(as.character(
  countyincome.tx$County.FIPS),

```



```

      3, side = "left", pad = "0")
names(countyincome.tx)[which(names(countyincome.tx) == "
  County.FIPS")] ← "COUNTYCODE"
names(countyincome.tx)[which(names(countyincome.tx) == "
  Median.Household.Income")] ← "MEDIANINCOME"
names(countyincome.tx)[which(names(countyincome.tx) == "
  Poverty.Percent.All.Ages")] ← "POVERTYPCT"
countyincome.tx$MEDIANINCOME ← as.numeric(gsub(",", "",
  as.character(countyincome.tx$MEDIANINCOME)))
countyincome.tx$POVERTYPCT ← as.numeric(gsub(",", "",
  as.character(countyincome.tx$POVERTYPCT)))
UrbanRural ← read.sas7bdat(str_c(dir, ":\users\scptex\dropbox
  \dissertation\data\NCHSUrbanRural\
  NCHSUrbruralcodes.sas7bdat"))
UrbanRural ← UrbanRural[which(UrbanRural$ST_ABBREV == "TX"), ]
UrbanRural$CO_FIPS ← str_pad(as.character(UrbanRural$CO_FIPS), 3,
  side = "left",
  pad = "0")
urbrurxref ← as.data.frame(cbind(c(1, 2, 3, 4, 5, 6), c("Large
  central metro",
  "Large fringe metro", "Medium metro", "Small metro", "
  Micropolitan", "Noncore")))
names(urbrurxref) ← c("CODE2006", "MetroStatus")
counties ← merge(counties, countyincome.tx, by = "COUNTYCODE")
counties ← merge(counties, UrbanRural, by.x = "COUNTYCODE", by.y
  = "CO_FIPS")

```

```

# *R Code Block* - loadzipfile - load file with zipcode level
# poverty level
# stats (ACS)
zippov ← read.csv(file = str_c(dir, ":\users\scptex\dropbox\
  dissertation\Data\ZIPPOV\zippovpct.csv"),
  head = TRUE, sep = ",") #read in zipcode povpct xref
names(zippov)[which(names(zippov) == "povpct")] ← "ZIPPOVPCT"
zippov$ZIPPOVPCT ← as.numeric(as.character(zippov$ZIPPOVPCT))

```

```

# *R Code Block* - makehospfile for hospital 2012 survey

hosp ← read.csv(file = str_c(dir, ":\users\scptex\Dropbox\
  Dissertation\Data\2012HospSurvey\HospSurv2012.csv"),
  head = TRUE, sep = ",")

hosp$EHR ← sub(2, 1, hosp$EHR)
hosp$EHR ← sub(3, 0, hosp$EHR)
hosp$EHR ← sub(4, 0, hosp$EHR)

```

```

hosp$EHR ← sub(5, 0, hosp$EHR)
hosp$EHR ← sub(6, 0, hosp$EHR)
hosp$EHR ← as.factor(hosp$EHR)

hosplink ← read.csv(file = str_c(dir, ":\users\scptex\Dropbox
  \Dissertation\Data\Hosp2012FromDSHS\HS12CD.csv"),
  head = TRUE, sep = ",")

hosplink2 ← read.csv(file = str_c(dir, ":\users\scptex\Dropbox
  \Dissertation\Data\Hosp2012FromDSHS\HS12DIR.csv"),
  head = TRUE, sep = ",")

hosp ← merge(hosp, hosplink, by = "FID")

hosp ← merge(hosp, hosplink2, by = "FID")

source(str_c(dir, ":\users\scptex\Dropbox\Dissertation\Text
  \code\counties\LoadCountyFiles.r"))

counties$Name ← toupper(counties$Name)
counties$Name ← gsub(" COUNTY", "", counties$Name)
counties$Name ← gsub(" ", "", counties$Name)
hosp$county ← gsub(" ", "", hosp$COUNTY)

hosp_comerge ← merge(hosp, counties, by.x = "COUNTY", by.y = "
  Name")

hosp_comerge$MEDIANINCOME ← as.numeric(gsub(",", "", as.character(
  hosp_comerge$MEDIANINCOME)))
hosp_comerge$POVERTYPCT ← as.numeric(gsub(",", "", as.character(
  hosp_comerge$POVERTYPCT)))

HRRZipXref ← read.csv(file = str_c(dir, ":\users\scptex\
  Dropbox\Dissertation\Data\ DartmouthHRRs\ZipHsaHrr11.csv"
  ),
  head = TRUE, sep = ",") #read in HRR-zip xref file

hosp_hrrmerge ← merge(hosp_comerge, HRRZipXref, by.x = "LOCOZIP",
  by.y = "zipcode11")

hosp_sub ← hosp_hrrmerge[, c("FID", "FacName", "EHR", "EHRYear",
  "D1A1", "D1F1",
  "D1G1", "D2A11", "D2B11", "D2C11", "D2D11", "D3E1", "D5B1", "
  MSTAT", "POPDENSITY",
  "MEDIANINCOME", "POVERTYPCT", "hrrnum")]

```

```

names(hosp_sub) ← c("fid", "facname", "ehr", "ehryear", "
  totalbeds", "inpatientdays",
  "edvisits", "medicaredischarges", "medicaredays", "
  medicaiddischarges",
  "medicaiddays", "totalrevenue", "charitycare", "metrostat", "
  popdensity",
  "medianincome", "povertypct", "hrrnum")

hosp_sub$size ← cut(hosp_sub$totalbeds, labels = c("Small", "
  Medium", "Large"),
  breaks = c(0, 100, 400, 5000))

hosp_sub$metrostat ← gsub("1", "metro", hosp_sub$metrostat)
hosp_sub$metrostat ← gsub("2", "nonmetro", hosp_sub$metrostat)
hosp_sub$povpct_lvls ← cut(hosp_sub$povertypct, 4)
hosp_sub$metrostat ← factor(hosp_sub$metrostat)

hosp_sub$mcdpct ← hosp_sub$medicaiddays/hosp_sub$inpatientdays
hosp_sub$mcdehrelig ← ifelse(hosp_sub$mcdpct > 0.1, 1, 0)

hosp_sub1 ← hosp_sub[hosp_sub$ehr == 0, ]
hosp_sub2 ← hosp_sub[hosp_sub$ehr == 1 & hosp_sub$ehryear ≥
  2009, ]
hosp_sub ← rbind(hosp_sub1, hosp_sub2)

hosp$hrrnum ← factor(hosp$hrrnum)

hospnew1 ← glm(ehr ~ size + hrrnum + survey_yr + size * survey_yr
  + hrrnum *
  survey_yr, data = hosp, family = "binomial")
hospnew2 ← glm(ehr ~ size + hrrnum + urbrurcode + survey_yr +
  size * survey_yr +
  hrrnum * survey_yr + urbrurcode * survey_yr, data = hosp,
  family = "binomial")
hospnew3 ← glm(ehr ~ size + hrrnum + urbrurcode + zippovpct_lvls
  + survey_yr +
  size * survey_yr + hrrnum * survey_yr + urbrurcode * survey_yr
  + zippovpct_lvls *
  survey_yr, data = hosp, family = "binomial")

stargazer(hospnew1, hospnew2, hospnew3, title = "Hospitals, Both
  Years, Factor Models with HRR, Table 1",
  no.space = TRUE, single.row = FALSE, digits = 2, label = "
  hosplogitnew_exp",

```

```

    keep = 1:10, align = TRUE, report = "vc*")

stargazer(hospnew1, hospnew2, hospnew3, title = "Hospitals, Both
  Years, Factor Models with HRR, Table 1",
  no.space = TRUE, single.row = FALSE, digits = 2, label = "
  hosplogitnew_exp",
  keep = 11:30, align = TRUE, report = "vc*")

stargazer(hospnew1, hospnew2, hospnew3, title = "Hospitals, Both
  Years, Factor Models with HRR, Table 1",
  no.space = TRUE, single.row = FALSE, digits = 2, label = "
  hosplogitnew_exp",
  keep = 31:50, align = TRUE, report = "vc*")

stargazer(hospnew1, hospnew2, hospnew3, title = "Hospitals, Both
  Years, Factor Models with HRR, Table 1",
  no.space = TRUE, single.row = FALSE, digits = 2, label = "
  hosplogitnew_exp",
  keep = 51:80, align = TRUE, report = "vc*")

# , covariate.labels=c('Size - Medium (100-400)', 'Size - Large (
  more than
# 400)', 'Metro-Large Fringe', 'Metro-Medium', 'Metro-Small', '
  Metro-Micro',
# 'Metro-Noncore', 'Poverty pct (10-20)', 'Poverty (20-30)', '
  Poverty pct
# (30-40)', 'Poverty pct (40-50)', 'Poverty pct (more than 50)')

# , covariate.labels=c('Survey Year', 'Size - Medium (100-400)*
  Year', 'Size -
# Large (more than 400)*Year', 'Metro-Large
# Fringe*Year', 'Metro-Medium*Year', 'Metro-Small*Year', '
  Metro-Micro*Year',
# 'Metro-Noncore*Year', 'Poverty pct (10-20)*Year', 'Poverty pct
# (20-30)*Year', 'Poverty pct (30-40)*Year', 'Poverty pct
# (40-50)*Year', 'Poverty pct (more than 50)*Year', 'Constant')

```

## K.9 R Code Used for Statistics, Tables, and Figures in Appendix D - Hospital EHR Adoption Estimated Using Linear Probability Models

```
# This document has been 'knit' together to include R code, R  
output,  
# narrative text, figures, and tables.
```

```
# *R Code Block* - loadlibraries - Load non-base R libraries  
rm(list = ls())  
library(stringr)  
library(xtable)  
library(ggplot2)  
library(plyr)  
library(corrgram)  
library(stargazer)  
library(BaylorEdPsych)  
library(sas7bdat)
```

```
# *R Code Block* - setdefaults - Set R defaults and constants
```

```
# options(width=50)
```

```
figtheme ← theme(plot.title = element_text(face = "bold", size =  
16)) + theme(axis.title.y = element_text(face = "bold",  
size = 14), axis.text.y = element_text(size = 12, color = "  
black"), axis.title.x = element_text(face = "bold",  
size = 14), axis.text.x = element_text(size = 12, color = "  
black"))
```

```
options(bitmapType = "cairo")
```

```
# *R Code Block* - printtbl - Load my function to print survey  
question  
# answer options for a particular question
```

```
PrintTbl ← function(CodeBook, SurveyField) {  
  
  c2 ← CodeBook[CodeBook$Name == SurveyField, c("Question", "  
Label", "Value")]  
  c2tab ← c2[c2$Question == 0, c("Label", "Value")]  
  c2lab ← c2[c2$Question == 1, "Label"]  
  tabout ← xtable(c2tab, label = SurveyField, caption = c2lab,  
align = "|l|X|l|")  
}
```

```

print((tabout), include.rownames = FALSE, tabular.environment
      = "tabularx",
      width = "\\textwidth", type = "latex")
}

printxtab ← function(col1, col2, col1names, col2names, lbl, cap,
  aln) {
  testttab ← table(col1, col2)
  attr(testttab, "dimnames") ← list(col1names, col2names)
  xtable(testttab, type = "latex", asis = "TRUE", floating =
    FALSE, label = lbl,
    caption = cap, align = aln)
}

# *R Code Block* - loadcountyfiles - codeblock to load countyxref
# file and
# rename county field
counties ← read.csv(file = str_c(dir, ":\users\\scptex\\dropbox
  \\dissertation\\data\\2011practitionersurvey_work\\
  countyxref.csv"),
  head = TRUE, sep = ",") #read in counties file
# names(counties)[2]='COUNTYCODE' #rename county number field in
# counties
counties$COUNTYCODE ← str_pad(as.character(counties$COUNTYCODE),
  3, side = "left",
  pad = "0")
counties$POPDENSITY ← as.numeric(gsub(",", "", as.character(
  counties$Population)))/as.numeric(gsub(",", "",
  "", as.character(counties$Size)))
countyincome ← read.csv(file = str_c(dir, ":\users\\scptex\\
  dropbox\\dissertation\\data\\est10ALL.csv"),
  head = TRUE, sep = ",", skip = 2) #read in county income file
#load countyincome file, limit it to TX counties and
#relevant columns, and format county FIPS field for
#matching to other files
countyincome.tx ← countyincome[countyincome$Postal == "TX", c("
  County.FIPS",
  "Name", "Poverty.Estimate.All.Ages", "Poverty.Percent.All.Ages",
  "Poverty.Estimate.Under.Age.18",
  "Poverty.Percent.Under.Age.18", "Poverty.Estimate.Ages.5.17",
  "Poverty.Percent.Ages.5.17",
  "Median.Household.Income", "Poverty.Estimate.Ages.0.4", "
  Poverty.Percent.Ages.0.4")]
countyincome.tx$County.FIPS ← str_pad(as.character(
  countyincome.tx$County.FIPS),

```

```

      3, side = "left", pad = "0")
names(countyincome.tx)[which(names(countyincome.tx) == "
  County.FIPS")] ← "COUNTYCODE"
names(countyincome.tx)[which(names(countyincome.tx) == "
  Median.Household.Income")] ← "MEDIANINCOME"
names(countyincome.tx)[which(names(countyincome.tx) == "
  Poverty.Percent.All.Ages")] ← "POVERTYPCT"
countyincome.tx$MEDIANINCOME ← as.numeric(gsub(",", "",
  as.character(countyincome.tx$MEDIANINCOME)))
countyincome.tx$POVERTYPCT ← as.numeric(gsub(",", "",
  as.character(countyincome.tx$POVERTYPCT)))
UrbanRural ← read.sas7bdat(str_c(dir, ":\users\scptex\dropbox
  \dissertation\data\NCHSUrbanRural\
  NCHSUrbruralcodes.sas7bdat"))
UrbanRural ← UrbanRural[which(UrbanRural$ST_ABBREV == "TX"), ]
UrbanRural$CO_FIPS ← str_pad(as.character(UrbanRural$CO_FIPS), 3,
  side = "left",
  pad = "0")
urbrurxref ← as.data.frame(cbind(c(1, 2, 3, 4, 5, 6), c("Large
  central metro",
  "Large fringe metro", "Medium metro", "Small metro", "
  Micropolitan", "Noncore")))
names(urbrurxref) ← c("CODE2006", "MetroStatus")
counties ← merge(counties, countyincome.tx, by = "COUNTYCODE")
counties ← merge(counties, UrbanRural, by.x = "COUNTYCODE", by.y
  = "CO_FIPS")

```

```

# *R Code Block* - loadzipfile - load file with zipcode level
# poverty level
# stats (ACS)
zippov ← read.csv(file = str_c(dir, ":\users\scptex\dropbox\
  dissertation\Data\ZIPPOV\zippovpct.csv"),
  head = TRUE, sep = ",") #read in zipcode povpct xref
names(zippov)[which(names(zippov) == "povpct")] ← "ZIPPOVPCT"
zippov$ZIPPOVPCT ← as.numeric(as.character(zippov$ZIPPOVPCT))

```

```

# *R Code Block* - makehospfile for hospital 2012 survey

hosp ← read.csv(file = str_c(dir, ":\users\scptex\Dropbox\
  Dissertation\Data\2012HospSurvey\HospSurv2012.csv"),
  head = TRUE, sep = ",")

hosp$EHR ← sub(2, 1, hosp$EHR)
hosp$EHR ← sub(3, 0, hosp$EHR)
hosp$EHR ← sub(4, 0, hosp$EHR)

```

```

hosp$EHR ← sub(5, 0, hosp$EHR)
hosp$EHR ← sub(6, 0, hosp$EHR)
hosp$EHR ← as.factor(hosp$EHR)

hosplink ← read.csv(file = str_c(dir, ":\users\scptex\Dropbox
  \Dissertation\Data\Hosp2012FromDSHS\HS12CD.csv"),
  head = TRUE, sep = ",")

hosplink2 ← read.csv(file = str_c(dir, ":\users\scptex\Dropbox
  \Dissertation\Data\Hosp2012FromDSHS\HS12DIR.csv"),
  head = TRUE, sep = ",")

hosp ← merge(hosp, hosplink, by = "FID")

hosp ← merge(hosp, hosplink2, by = "FID")

source(str_c(dir, ":\users\scptex\Dropbox\Dissertation\Text
  \code\counties\LoadCountyFiles.r"))

counties$Name ← toupper(counties$Name)
counties$Name ← gsub(" COUNTY", "", counties$Name)
counties$Name ← gsub(" ", "", counties$Name)
hosp$county ← gsub(" ", "", hosp$COUNTY)

hosp_comerge ← merge(hosp, counties, by.x = "COUNTY", by.y = "
  Name")

hosp_comerge$MEDIANINCOME ← as.numeric(gsub(",", "", as.character(
  hosp_comerge$MEDIANINCOME)))
hosp_comerge$POVERTYPCT ← as.numeric(gsub(",", "", as.character(
  hosp_comerge$POVERTYPCT)))

HRRZipXref ← read.csv(file = str_c(dir, ":\users\scptex\
  Dropbox\Dissertation\Data\ DartmouthHRRs\ZipHsaHrr11.csv"
  ),
  head = TRUE, sep = ",") #read in HRR-zip xref file

hosp_hrrmerge ← merge(hosp_comerge, HRRZipXref, by.x = "LOCOZIP",
  by.y = "zipcode11")

hosp_sub ← hosp_hrrmerge[, c("FID", "FacName", "EHR", "EHYear",
  "D1A1", "D1F1",
  "D1G1", "D2A11", "D2B11", "D2C11", "D2D11", "D3E1", "D5B1", "
  MSTAT", "POPDENSITY",
  "MEDIANINCOME", "POVERTYPCT", "hrrnum")]

```



```

names(hosp_sub) ← c("fid", "facname", "ehr", "ehryear", "
  totalbeds", "inpatientdays",
  "edvisits", "medicaredischarges", "medicaredays", "
  medicaiddischarges",
  "medicaiddays", "totalrevenue", "charitycare", "metrostat", "
  popdensity",
  "medianincome", "povertypct", "hrrnum")

hosp_sub$size ← cut(hosp_sub$totalbeds, labels = c("Small", "
  Medium", "Large"),
  breaks = c(0, 100, 400, 5000))

hosp_sub$metrostat ← gsub("1", "metro", hosp_sub$metrostat)
hosp_sub$metrostat ← gsub("2", "nonmetro", hosp_sub$metrostat)
hosp_sub$povpct_lvls ← cut(hosp_sub$povertypct, 4)
hosp_sub$metrostat ← factor(hosp_sub$metrostat)

hosp_sub$mcdpct ← hosp_sub$medicaiddays/hosp_sub$inpatientdays
hosp_sub$mcdehrelig ← ifelse(hosp_sub$mcdpct > 0.1, 1, 0)

hosp_sub1 ← hosp_sub[hosp_sub$ehr == 0, ]
hosp_sub2 ← hosp_sub[hosp_sub$ehr == 1 & hosp_sub$ehryear ≥
  2009, ]
hosp_sub ← rbind(hosp_sub1, hosp_sub2)

```

```

hosp$ehr ← as.numeric(hosp$ehr) - 1

hospnew1 ← glm(ehr ~ size + survey_yr + size * survey_yr, data =
  hosp, family = "gaussian")
hospnew2 ← glm(ehr ~ size + urbrurcode + survey_yr + size *
  survey_yr + urbrurcode *
  survey_yr, data = hosp, family = "gaussian")
hospnew3 ← glm(ehr ~ size + urbrurcode + zippovpct_lvls +
  survey_yr + size *
  survey_yr + urbrurcode * survey_yr + zippovpct_lvls *
  survey_yr, data = hosp,
  family = "gaussian")

```

```

stargazer(hospnew1, hospnew2, hospnew3, title = "Hospitals, Both
  Years, LPM Factor Models, Explanatory Variables",
  no.space = TRUE, single.row = FALSE, digits = 2, label = "
  hosplogitnew_exp",
  keep = 1:12, align = TRUE, covariate.labels = c("Size - Medium
  (100-400)",

```

```

    "Size - Large (more than 400)", "Metro-Large Fringe", "
      Metro-Medium",
    "Metro-Small", "Metro-Micro", "Metro-Noncore", "Poverty
      pct (10-20)",
    "Poverty (20-30)", "Poverty pct (30-40)", "Poverty pct (40
      -50)", "Poverty pct (more than 50)",
    report = "vc*")

stargazer(hospnew1, hospnew2, hospnew3, title = "Hospitals, Both
  Years, LPM Factor Models (cont), Year Effects",
  no.space = TRUE, single.row = FALSE, digits = 2, label = "
  hosplogitnew_yreff",
  keep = 13:30, align = TRUE, covariate.labels = c("Survey Year"
    , "Size - Medium (100-400)*Year",
    "Size - Large (more than 400)*Year", "Metro-Large Fringe*
      Year", "Metro-Medium*Year",
    "Metro-Small*Year", "Metro-Micro*Year", "Metro-Noncore*
      Year", "Poverty pct (10-20)*Year",
    "Poverty pct (20-30)*Year", "Poverty pct (30-40)*Year", "
      Poverty pct (40-50)*Year",
    "Poverty pct (more than 50)*Year", "Constant"), report = "
    vc*")

```

## K.10 R Code Used for Statistics, Tables, and Figures in Appendix E - Predicting Physician EHR Adoption Using Survival Analysis

```
# This document has been 'knit' together to include R code, R
  output,
# narrative text, figures, and tables.
```

```
# *R Code Block* - loadlibraries - Load non-base R libraries
rm(list = ls())
library(stringr)
library(xtable)
library(ggplot2)
library(plyr)
library(corrgram)
library(stargazer)
library(BaylorEdPsych)
library(sas7bdat)
```

```
# *R Code Block* - setdefaults - Set R defaults and constants
# options(width=50)

figtheme ← theme(plot.title = element_text(face = "bold", size =
  16)) + theme(axis.title.y = element_text(face = "bold",
  size = 14), axis.text.y = element_text(size = 12, color = "
  black"), axis.title.x = element_text(face = "bold",
  size = 14), axis.text.x = element_text(size = 12, color = "
  black"))

options(bitmapType = "cairo")
```

```
# *R Code Block* - initvalues1
# init vector for first 12 years of adoption level
adopt_pet ← rep(0, 3)
adopt_hi ← rep(0, 3)
adopt_lo ← rep(0, 3)

# populate years 10-12; the 3 years for which the 2006 paper has
  data
adopt_pet ← c(12.8, 14.4, 18.325)
confint ← c(0.75, 1.27, sd(c(17.1, 17.2, 18, 21)))
adopt_hi ← adopt_pet + confint
adopt_lo ← adopt_pet - confint
```

```

# *R Code Block* - initvalues2 initialize vector holding years
yr ← c(1:40)

# init starting values of p and q to estimated values from 2006
  paper
p_start ← 0.00539624624901535
q_start ← 0.167297806617814

# *R Code Block* - loadpract2012
source(str_c(dir, ":\users\\scptex\\Dropbox\\Dissertation\\Text
  \\code\\logit_practfile\\Pract2012DataLoad.r"))

library(survival)

pract.surv ← pract.sub[!(pract.sub$EHR5 %in% c(88, 99)), c("EHR1"
  , "EHR5")]

# recoding some values code EHR1 2 to 1; make paperless or mostly
  paperless
# 1 code EHR1 3, 4, and 5 to 0; make everyting else 0

pract.surv$EHR1 ← as.numeric(sub(1, 1, pract.surv$EHR1))
pract.surv$EHR1 ← as.numeric(sub(2, 1, pract.surv$EHR1))
pract.surv$EHR1 ← as.numeric(sub(3, 0, pract.surv$EHR1))
pract.surv$EHR1 ← as.numeric(sub(4, 0, pract.surv$EHR1))
pract.surv$EHR1 ← as.numeric(sub(5, 0, pract.surv$EHR1))

# set 1980 to time==0
pract.surv$time ← pract.surv$EHR5 - 1980

# pract.surv$EHR1 pract.surv$time pract.surv$EHR5

# code NAs in time to 32; these correspond to 0s in adoption
pract.surv$time[is.na(pract.surv$time)] ← 32

# table(pract.surv$EHR5)

psurv ← Surv(pract.surv$time, pract.surv$EHR1)

# pract.surv$censored←0

survout ← survfit(psurv ~ 1, conf.type = "none")

plot(survout)

```

```
# cumulative hazard function  
cumulhaz ← -log(survout$surv)
```

```
# plot of cumulative hazard against time  
plot(survout$time, cumulhaz)
```

## K.11 R Code Used for Statistics, Tables, and Figures in Appendix F - Bass Model Sensitivity Analysis by Effect Year

```
# This document has been 'knit' together to include R code, R
  output,
# narrative text, figures, and tables.
```

```
# *R Code Block* - loadlibraries - Load non-base R libraries
rm(list = ls())
library(stringr)
library(xtable)
library(ggplot2)
library(plyr)
library(corrgram)
library(stargazer)
library(BaylorEdPsych)
library(sas7bdat)
```

```
# *R Code Block* - setdefaults - Set R defaults and constants
# options(width=50)

figtheme ← theme(plot.title = element_text(face = "bold", size =
  16)) + theme(axis.title.y = element_text(face = "bold",
  size = 14), axis.text.y = element_text(size = 12, color = "
  black"), axis.title.x = element_text(face = "bold",
  size = 14), axis.text.x = element_text(size = 12, color = "
  black"))

options(bitmapType = "cairo")
```

```
# *R Code Block* - initvalues1
# init vector for first 12 years of adoption level
adopt_pet ← rep(0, 3)
adopt_hi ← rep(0, 3)
adopt_lo ← rep(0, 3)

# populate years 10-12; the 3 years for which the 2006 paper has
  data
adopt_pet ← c(12.8, 14.4, 18.325)
confint ← c(0.75, 1.27, sd(c(17.1, 17.2, 18, 21)))
adopt_hi ← adopt_pet + confint
adopt_lo ← adopt_pet - confint
```

```

# *R Code Block* - initvalues2 initialize vector holding years
yr <- c(1:40)

# init starting values of p and q to estimated values from 2006
  paper
p_start <- 0.00539624624901535
q_start <- 0.167297806617814

# *R Code Block* - bassmodelfunc bass diffusion model
predbass <- function(pfn_pandq, pfn_t) {
  (1 - exp(-(pfn_pandq[1] + pfn_pandq[2]) * pfn_t))/(1 + (
    pfn_pandq[2]/pfn_pandq[1]) *
    exp(-(pfn_pandq[1] + pfn_pandq[2]) * pfn_t)) * 100
}

# *R Code Block* - meandifffunc mean difference function to take
  the average
# diff between the observed adoption levels and the calculated
  adoption
# levels
meandiff <- function(mfn_pandq, mfn_func, mfn_obs, mfn_t,
  mfn_range) {
  mean(abs(mfn_obs - mfn_func(mfn_pandq, mfn_t)[mfn_range[1]:
    mfn_range[2]]))
}

# *R Code Block* - initpandq init p and q to close to their values
  in the
# spreadsheet
pandqout <- c(0.02, 0.05)

# *R Code Block* - runopt run optimization
bassopt <- optim(par = pandqout, fn = meandiff, mfn_func =
  predbass, mfn_obs = adopt_pct,
  mfn_t = yr, mfn_range = c(10, 12))

bassopt_hi <- optim(par = pandqout, fn = meandiff, mfn_func =
  predbass, mfn_obs = adopt_hi,
  mfn_t = yr, mfn_range = c(10, 12))

bassopt_lo <- optim(par = pandqout, fn = meandiff, mfn_func =
  predbass, mfn_obs = adopt_lo,
  mfn_t = yr, mfn_range = c(10, 12))

# *R Code Block* - loadpract2012

```

```

source(str_c(dir, ":\ users\sptex\Dropbox\Dissertation\Text
\\\\code\\logit_practfile\\Pract2012DataLoad.r"))

# *R Code Block* - buildayd
buildadoptyrdistro ← function(pd) {

  adoptyrdistro ← data.frame(1989:2012)
  names(adoptyrdistro) ← "Year"

  tmpadoptyrdistro ← as.data.frame(table(pd$EHR5[!(pd$EHR5 %in%
    c(88, 99))]))
  # names(b[1]) ← 'Year'

  adoptyrdistro ← merge(adoptyrdistro, tmpadoptyrdistro, by.x =
    "Year", by.y = "Var1",
    all = TRUE)

  adoptyrdistro[is.na(adoptyrdistro)] ← 0

  cumullv1 ← rep(0, nrow(adoptyrdistro))
  adoptyrdistro ← cbind(adoptyrdistro, cumullv1)

  numrows ← NROW(pd$EHR5[!(pd$EHR5 %in% c(88, 99))])

  for (i in 1:nrow(adoptyrdistro)) {
    adoptyrdistro$cumullv1[i] ← sum(adoptyrdistro$Freq[1:i])
  }

  cumulpct ← rep(0, nrow(adoptyrdistro))
  adoptyrdistro ← cbind(adoptyrdistro, cumulpct)

  # adjust for part-year 2012
  jun21 ← 173
  oct12 ← 286
  midsurv ← (jun21 + oct12)/2
  len12 ← 366

  portionleft ← (len12 - midsurv)/len12
  nextyrcount ← count(pd$EHR1[which(pd$EHR1 == 3)])$freq
  alloc ← portionleft * nextyrcount

  adoptyrdistro[which(adoptyrdistro$Year == 2012), ]$Freq ←
    adoptyrdistro[which(adoptyrdistro$Year ==
    2012), ]$Freq + round(alloc)
}

```



```

adoptyrdistro$scumulpet ← adoptyrdistro$scumullvl/numrows * 100

return(adoptyrdistro)
}

adoptyrdistro ← buildadoptyrdistro(pract.sub)

# *R Code Block* - loadcalcbassfunction
calcbass ← function(practdata, bassfunc, difffunc, yrlist,
  initpnq, aydfunc) {

  ad ← aydfunc(practdata)

  pnq_surv_pre ← optim(par = initpnq, fn = difffunc, mfn_func =
    bassfunc,
    mfn_obs = ad$scumulpet[4:20], mfn_t = yrlist, mfn_range = c
      (1, 17))
  pnq_surv_post ← optim(par = initpnq, fn = difffunc, mfn_func
    = bassfunc,
    mfn_obs = ad$scumulpet[4:24], mfn_t = yrlist, mfn_range = c
      (1, 21))
  return(cbind(pnq_surv_pre, pnq_surv_post))
}

# *R Code Block* - calcnewbass
a ← calcbass(pract.sub, predbass, meandiff, yr, pandqout,
  buildadoptyrdistro)

pnq_surv_pre ← a[1, ]$pnq_surv_pre
pnq_surv_post ← a[1, ]$pnq_surv_post

# *R Code Block* - bootstrap2 predicted adoption rates for 2011
-2014

# if
# (!file.exists('C:\\Users\\scptex\\Dropbox\\Dissertation\\Text\\
  Section3_RDD\\bootstrap.csv'))
# {
predvalues ← as.data.frame(matrix(ncol = 23, nrow = 10))
names(predvalues) ← as.character(1992:2014)

for (j in 1:1000) {

  practrep ← pract.sub[sample(nrow(pract.sub), nrow(pract.sub),
    replace = TRUE),

```

```

]

pnqprenpost ← calcbass(practrep, predbass, meandiff, yr,
  pandqout, buildadoptyrdistro)

predvalues[j, ] ← predbass(pnqprenpost[1, ]$pnq_surv_pre,
  1:23)
}
# write.csv(predvalues, file='c:\\users\\scptex\\dropbox\\
  dissertation\\text\\Section3_RDD\\bootstrap.csv')
# }

# if
# (file.exists('C:\\Users\\scptex\\Dropbox\\Dissertation\\Text\\
  Section3_RDD\\bootstrap.csv'))
# {
# predvalues ← read.csv(file='c:\\users\\scptex\\dropbox\\
  dissertation\\text\\Section3_RDD\\bootstrap.csv', header=TRUE,
  row.names=1)
names(predvalues) ← 1992:2014
# }

newconfint_lo ← apply(predvalues, 2, sort)[26, ]
newconfint_hi ← apply(predvalues, 2, sort)[975, ]
predpre ← as.data.frame(predbass(pnq_surv_pre, 1:23))
names(predpre) ← "Predicted"
row.names(predpre) ← 1992:2014

# *R Code Block* - basicconfintplots
ncil ← as.data.frame(newconfint_lo)
ncih ← as.data.frame(newconfint_hi)

ncih_df ← as.data.frame(ncih)
ncih_df$year ← as.numeric(row.names(ncih_df))

ncil_df ← as.data.frame(ncil)
ncil_df$year ← as.numeric(row.names(ncil_df))

predpre$year ← as.numeric(row.names(predpre))

confintplot08 ← ggplot() + geom_line(data = ncih_df, aes(x = year
  , y = newconfint_hi,
  group = 1), color = "blue", size = 1, linetype = 2) +
  geom_line(data = ncil_df,

```

```

aes(x = year, y = newconfint_lo, group = 1), color = "blue",
    size = 1, linetype = 2) +
geom_line(data = predpre, aes(x = year, y = Predicted, group =
    1), color = "blue",
    size = 1) + geom_point(data = adoptyrdistro[which(
    adoptyrdistro$Year %in%
    1992:2012), ], aes(x = Year, y = cumulpct), size = 3) +
    geom_vline(xintercept = 2008,
    linetype = "dotted")

```

```

# *R Code Block* - plothyptest08
hyptest <- adoptyrdistro[which(adoptyrdistro$Year %in% 2009:2012),
    c("cumulpct",
    "Year")]
names(hyptest) <- c("cumulpct", "year")

hyptest <- merge(hyptest, ncih_df, by = "year")
hyptest <- merge(hyptest, ncil_df, by = "year")
disphyptest <- hyptest[, c(1, 4, 2, 3)]
row.names(disphyptest) <- NULL

names(disphyptest) <- c("Year", "Conf. Int. , Low", "Measured
    Adoption", "Conf. Int. , High")

xdisphyptest08 <- xtable(disphyptest, caption = "Actual Adoption
    Levels and Confidence Intervals, 2008 Implementation",
    label = "tbl.hyptest", align = "|c|c|c|c|c|", type = "latex",
    asis = "TRUE",
    floating = FALSE)

# this is the predicted adoption level for 2014 under the adoption
# scenario;
# 1991==yr=0

```

```

# *R Code Block* - loadcalcbassfunction
calcbass <- function(practdata, bassfunc, difffunc, yrlist,
    initpnq, aydfunc) {

    ad <- aydfunc(practdata)

    pnq_surv_pre <- optim(par = initpnq, fn = difffunc, mfn_func =
    bassfunc,
    mfn_obs = ad$cumulpct[4:22], mfn_t = yrlist, mfn_range = c
    (1, 19))

```

```

    pnq_surv_post ← optim(par = initpnq, fn = diffunc, mfn_func
      = bassfunc,
      mfn_obs = ad$cumulpct[4:24], mfn_t = yrlist, mfn_range = c
        (1, 21))
    return(cbind(pnq_surv_pre, pnq_surv_post))
  }

# *R Code Block* - calcnewbass
a ← calcbass(pract.sub, predbass, meandiff, yr, pandqout,
  buildadoptyrdistro)

pnq_surv_pre ← a[1, ]$pnq_surv_pre
pnq_surv_post ← a[1, ]$pnq_surv_post

# *R Code Block* - bootstrap2 predicted adoption rates for 2011
-2014

# if
# (!file.exists('C:\\Users\\scptex\\Dropbox\\Dissertation\\Text\\
  Section3_RDD\\bootstrap.csv'))
# {
predvalues ← as.data.frame(matrix(ncol = 23, nrow = 10))
names(predvalues) ← as.character(1992:2014)

for (j in 1:1000) {

  practrep ← pract.sub[sample(nrow(pract.sub), nrow(pract.sub),
    replace = TRUE),
    ]

  pnqprenpost ← calcbass(practrep, predbass, meandiff, yr,
    pandqout, buildadoptyrdistro)

  predvalues[j, ] ← predbass(pnqprenpost[1, ]$pnq_surv_pre,
    1:23)

}
# write.csv(predvalues, file='c:\\users\\scptex\\dropbox\\
  dissertation\\text\\Section3_RDD\\bootstrap.csv')
# }

# if
# (file.exists('C:\\Users\\scptex\\Dropbox\\Dissertation\\Text\\
  Section3_RDD\\bootstrap.csv'))
# {

```

```

# predvalues←read.csv(file='c:\\users\\scptex\\dropbox\\
  dissertation\\text\\Section3_RDD\\bootstrap.csv',header=TRUE,
  row.names=1)
names(predvalues) ← 1992:2014
# }

newconfint_lo ← apply(predvalues, 2, sort)[26, ]
newconfint_hi ← apply(predvalues, 2, sort)[975, ]
predpre ← as.data.frame(predbass(pnq_surv_pre, 1:23))
names(predpre) ← "Predicted"
row.names(predpre) ← 1992:2014

# *R Code Block* - basicconfintplots
ncil ← as.data.frame(newconfint_lo)
ncih ← as.data.frame(newconfint_hi)

ncih_df ← as.data.frame(ncih)
ncih_df$year ← as.numeric(row.names(ncih_df))

ncil_df ← as.data.frame(ncil)
ncil_df$year ← as.numeric(row.names(ncil_df))

predpre$year ← as.numeric(row.names(predpre))

confintplot10 ← ggplot() + geom_line(data = ncih_df, aes(x = year
, y = newconfint_hi,
  group = 1), color = "blue", size = 1, linetype = 2) +
  geom_line(data = ncil_df,
  aes(x = year, y = newconfint_lo, group = 1), color = "blue",
  size = 1, linetype = 2) +
  geom_line(data = predpre, aes(x = year, y = Predicted, group =
  1), color = "blue",
  size = 1) + geom_point(data = adoptyrdistro[which(
  adoptyrdistro$Year %in%
  1992:2012), ], aes(x = Year, y = cumulpct), size = 3) +
  geom_vline(xintercept = 2010,
  linetype = "dotted")

# *R Code Block* - plothyptest10
hyptest ← adoptyrdistro[which(adoptyrdistro$Year %in% 2011:2012),
  c("cumulpct",
  "Year")]
names(hyptest) ← c("cumulpct", "year")

hyptest ← merge(hyptest, ncih_df, by = "year")

```

```

hypptest ← merge(hypptest, ncil_df, by = "year")
disphyptest ← hypptest[, c(1, 4, 2, 3)]
row.names(disphyptest) ← NULL

names(disphyptest) ← c("Year", "Conf. Int. , Low", "Measured
  Adoption", "Conf. Int. , High")

xdisphyptest10 ← xtable(disphyptest, caption = "Actual Adoption
  Levels and Confidence Intervals , 2010 Implementation",
  label = "tbl.hypptest", align = "|c|c|c|c|c|", type = "latex",
  asis = "TRUE",
  floating = FALSE)

# print(xdisphyptest10, include.rownames=FALSE)

# this is the predicted adoption level for 2014 under the adoption
  scenario;
# 1991==yr=0

# *R Code Block* - loadcalcbassfunction
calcbass ← function(practdata, bassfunc, difffunc, yrlist,
  initpnq, aydfunc) {

  ad ← aydfunc(practdata)

  pnq_surv_pre ← optim(par = initpnq, fn = difffunc, mfn_func =
    bassfunc,
    mfn_obs = ad$cumulpct[4:23], mfn_t = yrlist, mfn_range = c
    (1, 20))
  pnq_surv_post ← optim(par = initpnq, fn = difffunc, mfn_func
    = bassfunc,
    mfn_obs = ad$cumulpct[4:24], mfn_t = yrlist, mfn_range = c
    (1, 21))
  return(cbind(pnq_surv_pre, pnq_surv_post))
}

# *R Code Block* - calcnewbass
a ← calcbass(pract.sub, predbass, meandiff, yr, pandqout,
  buildadoptyrdistro)

pnq_surv_pre ← a[1, ]$pnq_surv_pre
pnq_surv_post ← a[1, ]$pnq_surv_post

# *R Code Block* - bootstrap2 predicted adoption rates for 2011
  -2014

```

```

# if
# (!file.exists('C:\\Users\\scptex\\Dropbox\\Dissertation\\Text\\
  Section3_RDD\\bootstrap.csv'))
# {
predvalues <- as.data.frame(matrix(ncol = 23, nrow = 10))
names(predvalues) <- as.character(1992:2014)

for (j in 1:1000) {

  practrep <- pract.sub[sample(nrow(pract.sub), nrow(pract.sub),
    replace = TRUE),
    ]

  pnqprenpost <- calcbass(practrep, predbass, meandiff, yr,
    pandqout, buildadoptyrdistro)

  predvalues[j, ] <- predbass(pnqprenpost[1, ]$pnq_surv_pre,
    1:23)

}
# write.csv(predvalues, file='c:\\users\\scptex\\dropbox\\
  dissertation\\text\\Section3_RDD\\bootstrap.csv')
# }

# if
# (file.exists('C:\\Users\\scptex\\Dropbox\\Dissertation\\Text\\
  Section3_RDD\\bootstrap.csv'))
# {
# predvalues<-read.csv(file='c:\\users\\scptex\\dropbox\\
  dissertation\\text\\Section3_RDD\\bootstrap.csv',header=TRUE,
  row.names=1)
names(predvalues) <- 1992:2014
# }

newconfint_lo <- apply(predvalues, 2, sort)[26, ]
newconfint_hi <- apply(predvalues, 2, sort)[975, ]
predpre <- as.data.frame(predbass(pnq_surv_pre, 1:23))
names(predpre) <- "Predicted"
row.names(predpre) <- 1992:2014

# *R Code Block* - basicconfintplots
ncil <- as.data.frame(newconfint_lo)
ncih <- as.data.frame(newconfint_hi)

```

```

ncih_df ← as.data.frame(ncih)
ncih_df$year ← as.numeric(row.names(ncih_df))

ncil_df ← as.data.frame(ncil)
ncil_df$year ← as.numeric(row.names(ncil_df))

predpre$year ← as.numeric(row.names(predpre))

confintplot11 ← ggplot() + geom_line(data = ncih_df, aes(x = year
, y = newconfint_hi,
group = 1), color = "blue", size = 1, linetype = 2) +
  geom_line(data = ncil_df,
aes(x = year, y = newconfint_lo, group = 1), color = "blue",
size = 1, linetype = 2) +
  geom_line(data = predpre, aes(x = year, y = Predicted, group =
1), color = "blue",
size = 1) + geom_point(data = adoptyrdistro[which(
adoptyrdistro$Year %in%
1992:2012), ], aes(x = Year, y = cumulpct), size = 3) +
  geom_vline(xintercept = 2011,
linetype = "dotted")

```

```

# *R Code Block* - plothyptest11
hyptest ← adoptyrdistro[which(adoptyrdistro$Year %in% 2012), c("
cumulpct",
"Year")]
names(hyptest) ← c("cumulpct", "year")

hyptest ← merge(hyptest, ncih_df, by = "year")
hyptest ← merge(hyptest, ncil_df, by = "year")
disphyptest ← hyptest[, c(1, 4, 2, 3)]
row.names(disphyptest) ← NULL

names(disphyptest) ← c("Year", "Conf. Int., Low", "Measured
Adoption", "Conf. Int., High")

xdisphyptest11 ← xtable(disphyptest, caption = "Actual Adoption
Levels and Confidence Intervals, 2011 Implementation",
label = "tbl.hyptest", align = "|c|c|c|c|c|", type = "latex",
asis = "TRUE",
floating = FALSE)

# print(xdisphyptest11, include.rownames=FALSE)

# this is the predicted adoption level for 2014 under the adoption

```



```
    scenario;  
# 1991==yr=0
```

```
confintplot08
```

```
print(xdisphyptest08 , include.rownames = FALSE)
```

```
confintplot10
```

```
print(xdisphyptest10 , include.rownames = FALSE)
```

```
confintplot11
```

```
print(xdisphyptest11 , include.rownames = FALSE)
```

## K.12 R Code Used for Statistics, Tables, and Figures in Appendix G - Robustness of Physician RD Design Models to Addition of Covariates

```
# *R Code Block* - loadlibraries - Load non-base R libraries
rm(list = ls())
library(stringr)
library(xtable)
library(ggplot2)
library(plyr)
library(corrgram)
library(stargazer)
library(BaylorEdPsych)
library(sas7bdat)
```

```
# *R Code Block* - init
figtheme ← theme(plot.title = element_text(face = "bold", size =
  16)) + theme(axis.title.y = element_text(face = "bold",
  size = 14), axis.text.y = element_text(size = 12, color = "
  black"), axis.title.x = element_text(face = "bold",
  size = 14), axis.text.x = element_text(size = 12, color = "
  black"))

dir ← "C"
```

```
# *R Code Block* - recrddprocess
source(str_c(dir, ":\users\sctex\Dropbox\Dissertation\Text
  \code\logit_practfile\Pract2012DataLoad.r"))

pract.sub$tenorfewer ← cut(pract.sub$DI4, breaks = c(0, 11, 1000)
  , labels = c(1,
  0))
pract.sub$recelig ← pract.sub$tenorfewer == 1 & pract.sub$DI7 ==
  1

names(pract.sub)[3] ← "COUNTYCODE"
pract.sub$COUNTYCODE ← as.character(pract.sub$COUNTYCODE)
pract.sub$COUNTYCODE ← str_pad(pract.sub$COUNTYCODE, 3, side = "
  left", pad = "0")

source(str_c(dir, ":\users\sctex\Dropbox\Dissertation\Text
  \code\Counties\LoadCountyFiles.r"))

pract.sub ← merge(pract.sub, counties, by = "COUNTYCODE")
```

```

pract.sub$Population ← as.numeric(gsub(",", "", as.character(
  pract.sub$Population)))
pract.sub$Size ← as.numeric(gsub(",", "", as.character(pract.sub$
  Size)))
pract.sub$Population ← as.numeric(pract.sub$Population)
pract.sub$Size ← as.numeric(pract.sub$Size)
pract.sub$POPDENSITY ← pract.sub$Population/pract.sub$Size

# to sub zeroes for 2s in coding for DI2, Medicaid =1, no=0 after
  sub
pract.sub$DI2 ← as.integer(sub(2, 0, pract.sub$DI2))

names(pract.sub)[which(names(pract.sub) == "DI2")] ← "MEDICAID"
names(pract.sub)[which(names(pract.sub) == "DI4")] ← "
  PRACTITIONERCOUNT"

pract.sub$EHR1 ← sub(2, 1, pract.sub$EHR1)
pract.sub$EHR1 ← sub(3, 0, pract.sub$EHR1)
pract.sub$EHR1 ← sub(4, 0, pract.sub$EHR1)
pract.sub$EHR1 ← sub(5, 0, pract.sub$EHR1)

pract.sub$DI7 ← sub(2, 0, pract.sub$DI7)
names(pract.sub)[which(names(pract.sub) == "DI7")] ← "PRIMARYCARE
  "
pract.sub$PRIMARYCARE ← as.factor(pract.sub$PRIMARYCARE)

names(pract.sub)[which(names(pract.sub) == "EHR1")] ← "EHR"
pract.sub$EHR ← as.numeric(pract.sub$EHR)
pract.sub$EHR ← as.factor(pract.sub$EHR)

pract.sub$MEDICAID ← factor(pract.sub$MEDICAID)
pract.sub$CODE2006 ← factor(pract.sub$CODE2006)

pract.sub ← merge(pract.sub, urbrurxref, by = "CODE2006")

pract.sub$MetroStatus ← factor(pract.sub$MetroStatus, levels(
  pract.sub$MetroStatus)[c(5,
    4, 6, 3, 2, 1)])
pract.sub$popdens_lvls ← as.ordered(cut(pract.sub$POPDENSITY, 4))
pract.sub$practent_lvls ← cut(pract.sub$PRACTITIONERCOUNT, breaks
  = c(0, 1,
    5, 10, 100, 1000), labels = c("Solo", "Small", "Medium", "
    Large", "Very Large"))

pract.sub$medinc_lvls ← cut(pract.sub$MEDIANINCOME, 4)

```

```

pract.sub$povpct_lvls ← cut(pract.sub$POVERTYPCT, 4)

pract.sub ← pract.sub[!(pract.sub$PRIMARYCARE %in% c(88, 99)), ]

pract.sub1 ← pract.sub[pract.sub$EHR == 0 & pract.sub$PRIMARYCARE
  == 1, ]
pract.sub2 ← pract.sub[pract.sub$EHR == 1 & pract.sub$PRIMARYCARE
  == 1 & pract.sub$EHR5 ≥
  2009, ]
pract.sub ← rbind(pract.sub1, pract.sub2)

pract.sub$EHR ← as.numeric(as.character(pract.sub$EHR))

recrddtest ← RDestimate(EHR ~ PRACTITIONERCOUNT, data = pract.sub
  , cutpoint = 10)

# Imbens-Kalyanaraman Optimal Bandwidth Calculation
recIKbwtst ← IKbandwidth(pract.sub$PRACTITIONERCOUNT, pract.sub$
  EHR, cutpoint = 10)

recrddlogit1 ← glm(EHR ~ MEDICAID + PRACTITIONERCOUNT + recelig ,
  data = pract.sub ,
  family = "binomial")
recrddlogit2 ← glm(EHR ~ MEDICAID + PRACTITIONERCOUNT + recelig +
  povpct_lvls ,
  data = pract.sub , family = "binomial")
recrddlogit3 ← glm(EHR ~ MEDICAID + PRACTITIONERCOUNT + recelig +
  povpct_lvls +
  MetroStatus, data = pract.sub , family = "binomial")

# *R Code Block* - calcrecpolyplotscov

rec_lin_bw5 ← lm(EHR ~ PRACTITIONERCOUNT + recelig + povpct_lvls
  + MetroStatus +
  PRACTITIONERCOUNT * recelig , data = pract.sub[which(pract.sub$
  PRACTITIONERCOUNT <
  15 & pract.sub$PRACTITIONERCOUNT > 5), ])

rec_lin_bw10 ← lm(EHR ~ PRACTITIONERCOUNT + recelig + povpct_lvls
  + MetroStatus +
  PRACTITIONERCOUNT * recelig , data = pract.sub[which(pract.sub$
  PRACTITIONERCOUNT <
  20), ])

rec_quad ← lm(EHR ~ PRACTITIONERCOUNT + I(PRACTITIONERCOUNT^2) +

```

```

recelig +
  povpct_lvls + MetroStatus + PRACTITIONERCOUNT * recelig + I(
    PRACTITIONERCOUNT^2) *
  recelig, data = pract.sub[which(pract.sub$PRACTITIONERCOUNT <
    30), ])

rec_cub ← lm(EHR ~ PRACTITIONERCOUNT + I(PRACTITIONERCOUNT^2) + I(
  PRACTITIONERCOUNT^3) +
  recelig + povpct_lvls + MetroStatus + PRACTITIONERCOUNT *
  recelig + I(PRACTITIONERCOUNT^2) *
  recelig + I(PRACTITIONERCOUNT^3) * recelig, data = pract.sub[
  which(pract.sub$PRACTITIONERCOUNT <
  30), ])

rec_quart ← lm(EHR ~ PRACTITIONERCOUNT + I(PRACTITIONERCOUNT^2) +
  I(PRACTITIONERCOUNT^3) +
  I(PRACTITIONERCOUNT^4) + recelig + povpct_lvls + MetroStatus +
  PRACTITIONERCOUNT *
  recelig + I(PRACTITIONERCOUNT^2) * recelig + I(
  PRACTITIONERCOUNT^3) * recelig +
  I(PRACTITIONERCOUNT^4) * recelig, data = pract.sub[which(
  pract.sub$PRACTITIONERCOUNT <
  30), ])

rec_disptbl_pt1 ← data.frame(x = seq(0, 10, by = 0.01), y1_5 = 0,
  y1_10 = 0,
  y2 = 0, y3 = 0, y4 = 0)
rec_disptbl_pt2 ← data.frame(x = seq(10, 20, by = 0.01), y1_5 =
  0, y1_10 = 0,
  y2 = 0, y3 = 0, y4 = 0)

coefs_rec_lin_bw5 ← coefficients(rec_lin_bw5)
coefs_rec_lin_bw10 ← coefficients(rec_lin_bw10)
coefs_rec_quad ← coefficients(rec_quad)
coefs_rec_cub ← coefficients(rec_cub)
coefs_rec_quart ← coefficients(rec_quart)

# calculate linear bw=5 estimates
rec_disptbl_pt1$y1_5 ← coefs_rec_lin_bw5[1] + coefs_rec_lin_bw5
[2] * rec_disptbl_pt1$x +
  (coefs_rec_lin_bw5[2] + coefs_rec_lin_bw5[4]) *
  rec_disptbl_pt1$x + coefs_rec_lin_bw5[3]
rec_disptbl_pt2$y1_5 ← coefs_rec_lin_bw5[1] + coefs_rec_lin_bw5
[2] * rec_disptbl_pt2$x

```

```

# calculate linear bw=10 estimates
rec_disptbl_pt1$y1_10 ← coefs_rec_lin_bw10 [1] + (
  coefs_rec_lin_bw10 [2] + coefs_rec_lin_bw10 [4]) *
  rec_disptbl_pt1$x + coefs_rec_lin_bw10 [3]
rec_disptbl_pt2$y1_10 ← coefs_rec_lin_bw10 [1] +
  coefs_rec_lin_bw10 [2] * rec_disptbl_pt2$x

# calculate quadratic estimates
rec_disptbl_pt1$y2 ← coefs_rec_quad [1] + (coefs_rec_quad [2] +
  coefs_rec_quad [5]) *
  rec_disptbl_pt1$x + (coefs_rec_quad [3] + coefs_rec_quad [6]) *
  I(rec_disptbl_pt1$x^2) +
  coefs_rec_quad [4]
rec_disptbl_pt2$y2 ← coefs_rec_quad [1] + coefs_rec_quad [2] *
  rec_disptbl_pt2$x +
  coefs_rec_quad [3] * I(rec_disptbl_pt2$x^2)

# calculate cubic estimates
rec_disptbl_pt1$y3 ← coefs_rec_cub [1] + (coefs_rec_cub [2] +
  coefs_rec_cub [6]) *
  rec_disptbl_pt1$x + (coefs_rec_cub [3] + coefs_rec_cub [7]) * I(
  rec_disptbl_pt1$x^2) +
  (coefs_rec_cub [4] + coefs_rec_cub [8]) * I(rec_disptbl_pt1$x^3)
  + coefs_rec_cub [5]
rec_disptbl_pt2$y3 ← coefs_rec_cub [1] + coefs_rec_cub [2] *
  rec_disptbl_pt2$x +
  coefs_rec_cub [3] * I(rec_disptbl_pt2$x^2) + coefs_rec_cub [4] *
  I(rec_disptbl_pt2$x^3)

# calculate quartic estimates
rec_disptbl_pt1$y4 ← coefs_rec_quart [1] + (coefs_rec_quart [2] +
  coefs_rec_quart [7]) *
  rec_disptbl_pt1$x + (coefs_rec_quart [3] + coefs_rec_quart [8])
  * I(rec_disptbl_pt1$x^2) +
  (coefs_rec_quart [4] + coefs_rec_quart [9]) * I(rec_disptbl_pt1$x
  x^3) + (coefs_rec_quart [5] +
  coefs_rec_quart [10]) * I(rec_disptbl_pt1$x^4) +
  coefs_rec_quart [6]

rec_disptbl_pt2$y4 ← coefs_rec_quart [1] + coefs_rec_quart [2] *
  rec_disptbl_pt2$x +
  coefs_rec_quart [3] * I(rec_disptbl_pt2$x^2) + coefs_rec_quart
  [4] * I(rec_disptbl_pt2$x^3) +
  coefs_rec_quart [5] * I(rec_disptbl_pt2$x^4)

```

```

# *R Code Block* - disprecpolyplotcov
ggplot() + geom_line(data = rec_disptbl_pt1, aes(x = x, y = y1_5,
  color = "Linear, bw=5")) +
  geom_line(data = rec_disptbl_pt2, aes(x = x, y = y1_5, color =
    "Linear, bw=5")) +
  geom_line(data = rec_disptbl_pt1, aes(x = x, y = y1_10, color
    = "Linear, bw=10")) +
  geom_line(data = rec_disptbl_pt2, aes(x = x, y = y1_10, color
    = "Linear, bw=10")) +
  geom_smooth(data = rec_disptbl_pt1, aes(x = x, y = y2, color =
    "Quadratic"),
    method = "lm", formula = y ~ x + I(x^2)) + geom_smooth(
    data = rec_disptbl_pt2,
  aes(x = x, y = y2, color = "Quadratic"), method = "lm",
    formula = y ~ x +
      I(x^2)) + geom_smooth(data = rec_disptbl_pt1, aes(x = x, y
    = y3, color = "Cubic"),
  method = "lm", formula = y ~ x + I(x^2) + I(x^3)) +
  geom_smooth(data = rec_disptbl_pt2,
  aes(x = x, y = y3, color = "Cubic"), method = "lm", formula =
    y ~ x + I(x^2) +
      I(x^3)) + geom_smooth(data = rec_disptbl_pt1, aes(x = x, y
    = y4, color = "Quartic"),
  method = "lm", formula = y ~ x + I(x^2) + I(x^3) + I(x^4)) +
  geom_smooth(data = rec_disptbl_pt2,
  aes(x = x, y = y4, color = "Quartic"), method = "lm", formula
    = y ~ x +
      I(x^2) + I(x^3) + I(x^4)) + geom_vline(xintercept = 10,
    linetype = "dotted") +
  geom_rect(mapping = aes(xmin = 8, xmax = 12, ymin = 0, ymax =
    1.2), fill = NA,
    color = "black", linetype = "dotted") +
    scale_color_discrete(name = "Model",
  breaks = c("Linear, bw=5", "Linear, bw=10", "Quadratic", "
    Cubic", "Quartic"))

```

```

# *R Code Block* - disprecpolyplotzoomcov
ggplot() + geom_line(data = rec_disptbl_pt1, aes(x = x, y = y1_5,
  color = "Linear, bw=5")) +
  geom_line(data = rec_disptbl_pt2, aes(x = x, y = y1_5, color =
    "Linear, bw=5")) +
  geom_line(data = rec_disptbl_pt1, aes(x = x, y = y1_10, color
    = "Linear, bw=10")) +
  geom_line(data = rec_disptbl_pt2, aes(x = x, y = y1_10, color
    = "Linear, bw=10")) +

```

```

geom_smooth(data = rec_disptbl_pt1, aes(x = x, y = y2, color =
  "Quadratic"),
  method = "lm", formula = y ~ x + I(x^2)) + geom_smooth(
  data = rec_disptbl_pt2,
  aes(x = x, y = y2, color = "Quadratic"), method = "lm",
  formula = y ~ x +
  I(x^2)) + geom_smooth(data = rec_disptbl_pt1, aes(x = x, y
  = y3, color = "Cubic"),
  method = "lm", formula = y ~ x + I(x^2) + I(x^3)) +
  geom_smooth(data = rec_disptbl_pt2,
  aes(x = x, y = y3, color = "Cubic"), method = "lm", formula =
  y ~ x + I(x^2) +
  I(x^3)) + geom_smooth(data = rec_disptbl_pt1, aes(x = x, y
  = y4, color = "Quartic"),
  method = "lm", formula = y ~ x + I(x^2) + I(x^3) + I(x^4)) +
  geom_smooth(data = rec_disptbl_pt2,
  aes(x = x, y = y4, color = "Quartic"), method = "lm", formula
  = y ~ x +
  I(x^2) + I(x^3) + I(x^4)) + geom_vline(xintercept = 10,
  linetype = "dotted") +
  xlim(8, 12) + ylim(0, 1.2) + scale_color_discrete(name = "
  Model", breaks = c("Linear", bw=5",
  "Linear", bw=10", "Quadratic", "Cubic", "Quartic"))

```

```

# *R Code Block* - recefftblcov

```

```

rec_eff_bw5 <- summary(rec_lin_bw5)$coefficients[3, c(1, 4)]
rec_eff_bw10 <- summary(rec_lin_bw10)$coefficients[3, c(1, 4)]
rec_eff_quad <- summary(rec_quad)$coefficients[4, c(1, 4)]
rec_eff_cub <- summary(rec_cub)$coefficients[5, c(1, 4)]
rec_eff_quart <- summary(rec_quart)$coefficients[6, c(1, 4)]

rec_eff <- rbind(rec_eff_bw5, rec_eff_bw10, rec_eff_quad,
  rec_eff_cub, rec_eff_quart)
rownames(rec_eff) <- c("Local linear, bandwidth=5", "Local linear,
  bandwidth=10",
  "Quadratic", "Cubic", "Quartic")

xtable(rec_eff, type = "latex", align = "|r|r|r|", asis = "TRUE",
  floating = FALSE,
  digits = 4, label = "rec_eff", caption = "REC RDD Effect Sizes
  at Cut-off with Covariates")

```



## K.13 R Code Used for Statistics, Tables, and Figures in Appendix H - Robustness of Physician RD Design Models to Different Bandwidth Calculations

```
# *R Code Block* - loadlibraries - Load non-base R libraries
rm(list = ls())
library(stringr)
library(xtable)
library(ggplot2)
library(plyr)
library(corrgram)
library(stargazer)
library(BaylorEdPsych)
library(sas7bdat)

# *R Code Block* - init
figtheme ← theme(plot.title = element_text(face = "bold", size =
  16)) + theme(axis.title.y = element_text(face = "bold",
  size = 14), axis.text.y = element_text(size = 12, color = "
  black"), axis.title.x = element_text(face = "bold",
  size = 14), axis.text.x = element_text(size = 12, color = "
  black"))

dir ← "C"

# *R Code Block* - recrddprocess
source(str_c(dir, ":\users\scptex\Dropbox\Dissertation\Text
  \code\logit_practfile\Pract2012DataLoad.r"))

pract.sub$tenorfewer ← cut(pract.sub$DI4, breaks = c(0, 11, 1000)
  , labels = c(1,
  0))
pract.sub$recelig ← pract.sub$tenorfewer == 1 & pract.sub$DI7 ==
  1

names(pract.sub)[3] ← "COUNTYCODE"
pract.sub$COUNTYCODE ← as.character(pract.sub$COUNTYCODE)
pract.sub$COUNTYCODE ← str_pad(pract.sub$COUNTYCODE, 3, side = "
  left", pad = "0")

source(str_c(dir, ":\users\scptex\Dropbox\Dissertation\Text
  \code\Counties\LoadCountyFiles.r"))

pract.sub ← merge(pract.sub, counties, by = "COUNTYCODE")
```

```

pract.sub$Population ← as.numeric(gsub(",", "", as.character(
  pract.sub$Population)))
pract.sub$Size ← as.numeric(gsub(",", "", as.character(pract.sub$
  Size)))
pract.sub$Population ← as.numeric(pract.sub$Population)
pract.sub$Size ← as.numeric(pract.sub$Size)
pract.sub$POPDENSITY ← pract.sub$Population/pract.sub$Size

# to sub zeroes for 2s in coding for DI2, Medicaid =1, no=0 after
  sub
pract.sub$DI2 ← as.integer(sub(2, 0, pract.sub$DI2))

names(pract.sub)[which(names(pract.sub) == "DI2")] ← "MEDICAID"
names(pract.sub)[which(names(pract.sub) == "DI4")] ← "
  PRACTITIONERCOUNT"

pract.sub$EHR1 ← sub(2, 1, pract.sub$EHR1)
pract.sub$EHR1 ← sub(3, 0, pract.sub$EHR1)
pract.sub$EHR1 ← sub(4, 0, pract.sub$EHR1)
pract.sub$EHR1 ← sub(5, 0, pract.sub$EHR1)

pract.sub$DI7 ← sub(2, 0, pract.sub$DI7)
names(pract.sub)[which(names(pract.sub) == "DI7")] ← "PRIMARYCARE
  "
pract.sub$PRIMARYCARE ← as.factor(pract.sub$PRIMARYCARE)

names(pract.sub)[which(names(pract.sub) == "EHR1")] ← "EHR"
pract.sub$EHR ← as.numeric(pract.sub$EHR)
pract.sub$EHR ← as.factor(pract.sub$EHR)

pract.sub$MEDICAID ← factor(pract.sub$MEDICAID)
pract.sub$CODE2006 ← factor(pract.sub$CODE2006)

pract.sub ← merge(pract.sub, urbrurxref, by = "CODE2006")

pract.sub$MetroStatus ← factor(pract.sub$MetroStatus, levels(
  pract.sub$MetroStatus)[c(5,
    4, 6, 3, 2, 1)])
pract.sub$popdens_lvls ← as.ordered(cut(pract.sub$POPDENSITY, 4))
pract.sub$practent_lvls ← cut(pract.sub$PRACTITIONERCOUNT, breaks
  = c(0, 1,
    5, 10, 100, 1000), labels = c("Solo", "Small", "Medium", "
    Large", "Very Large"))

pract.sub$medinc_lvls ← cut(pract.sub$MEDIANINCOME, 4)

```

```

pract.sub$povpct_lvls ← cut(pract.sub$POVERTYPCT, 4)

pract.sub ← pract.sub[!(pract.sub$PRIMARYCARE %in% c(88, 99)), ]

pract.sub1 ← pract.sub[pract.sub$EHR == 0 & pract.sub$PRIMARYCARE
  == 1, ]
pract.sub2 ← pract.sub[pract.sub$EHR == 1 & pract.sub$PRIMARYCARE
  == 1 & pract.sub$EHR5 ≥
  2009, ]
pract.sub ← rbind(pract.sub1, pract.sub2)

pract.sub$EHR ← as.numeric(as.character(pract.sub$EHR))

recrddtest ← RDestimate(EHR ~ PRACTITIONERCOUNT, data = pract.sub
  , cutpoint = 10)

# Imbens-Kalyanaraman Optimal Bandwidth Calculation
recIKbwtst ← IKbandwidth(pract.sub$PRACTITIONERCOUNT, pract.sub$
  EHR, cutpoint = 10)

recrddlogit1 ← glm(EHR ~ MEDICAID + PRACTITIONERCOUNT + recelig,
  data = pract.sub,
  family = "binomial")
recrddlogit2 ← glm(EHR ~ MEDICAID + PRACTITIONERCOUNT + recelig +
  povpct_lvls,
  data = pract.sub, family = "binomial")
recrddlogit3 ← glm(EHR ~ MEDICAID + PRACTITIONERCOUNT + recelig +
  povpct_lvls +
  MetroStatus, data = pract.sub, family = "binomial")

# *R Code Block* - calcrecpolyplots_bw

rec_lin_bw1 ← lm(EHR ~ PRACTITIONERCOUNT + recelig +
  PRACTITIONERCOUNT * recelig,
  data = pract.sub[which(pract.sub$PRACTITIONERCOUNT ≤ 11 &
  pract.sub$PRACTITIONERCOUNT ≥
  9), ])

rec_lin_bw3 ← lm(EHR ~ PRACTITIONERCOUNT + recelig +
  PRACTITIONERCOUNT * recelig,
  data = pract.sub[which(pract.sub$PRACTITIONERCOUNT ≤ 13 &
  pract.sub$PRACTITIONERCOUNT ≥
  7), ])

rec_lin_bw5 ← lm(EHR ~ PRACTITIONERCOUNT + recelig +

```

```

PRACTITIONERCOUNT * recelig ,
  data = pract.sub [which (pract.sub$PRACTITIONERCOUNT ≤ 15 &
    pract.sub$PRACTITIONERCOUNT ≥
    5), ])

rec_lin_bw10 ← lm(EHR ~ PRACTITIONERCOUNT + recelig +
  PRACTITIONERCOUNT * recelig ,
  data = pract.sub [which (pract.sub$PRACTITIONERCOUNT ≤ 20), ])

rec_quad ← lm(EHR ~ PRACTITIONERCOUNT + I (PRACTITIONERCOUNT^2) +
  recelig +
  PRACTITIONERCOUNT * recelig + I (PRACTITIONERCOUNT^2) * recelig
  , data = pract.sub [which (pract.sub$PRACTITIONERCOUNT ≤
  30), ])

rec_quad5 ← lm(EHR ~ PRACTITIONERCOUNT + I (PRACTITIONERCOUNT^2) +
  recelig +
  PRACTITIONERCOUNT * recelig + I (PRACTITIONERCOUNT^2) * recelig
  , data = pract.sub [which (pract.sub$PRACTITIONERCOUNT ≤
  15 & pract.sub$PRACTITIONERCOUNT ≥ 5), ])

rec_quad10 ← lm(EHR ~ PRACTITIONERCOUNT + I (PRACTITIONERCOUNT^2)
  + recelig +
  PRACTITIONERCOUNT * recelig + I (PRACTITIONERCOUNT^2) * recelig
  , data = pract.sub [which (pract.sub$PRACTITIONERCOUNT ≤
  20), ])

rec_cub ← lm(EHR ~ PRACTITIONERCOUNT + I (PRACTITIONERCOUNT^2) + I
  (PRACTITIONERCOUNT^3) +
  recelig + PRACTITIONERCOUNT * recelig + I (PRACTITIONERCOUNT^2)
  * recelig +
  I (PRACTITIONERCOUNT^3) * recelig , data = pract.sub [which (
  pract.sub$PRACTITIONERCOUNT ≤
  30), ])

rec_cub5 ← lm(EHR ~ PRACTITIONERCOUNT + I (PRACTITIONERCOUNT^2) +
  I (PRACTITIONERCOUNT^3) +
  recelig + PRACTITIONERCOUNT * recelig + I (PRACTITIONERCOUNT^2)
  * recelig +
  I (PRACTITIONERCOUNT^3) * recelig , data = pract.sub [which (
  pract.sub$PRACTITIONERCOUNT ≤
  15 & pract.sub$PRACTITIONERCOUNT ≥ 5), ])

rec_cub10 ← lm(EHR ~ PRACTITIONERCOUNT + I (PRACTITIONERCOUNT^2) +
  I (PRACTITIONERCOUNT^3) +

```

```

recelig + PRACTITIONERCOUNT * recelig + I(PRACTITIONERCOUNT^2)
  * recelig +
I(PRACTITIONERCOUNT^3) * recelig, data = pract.sub[which(
  pract.sub$PRACTITIONERCOUNT ≤
20), ])

rec_quart ← lm(EHR ~ PRACTITIONERCOUNT + I(PRACTITIONERCOUNT^2) +
  I(PRACTITIONERCOUNT^3) +
  I(PRACTITIONERCOUNT^4) + recelig + PRACTITIONERCOUNT * recelig
  + I(PRACTITIONERCOUNT^2) *
  recelig + I(PRACTITIONERCOUNT^3) * recelig + I(
  PRACTITIONERCOUNT^4) * recelig,
  data = pract.sub[which(pract.sub$PRACTITIONERCOUNT ≤ 30), ])

rec_quart5 ← lm(EHR ~ PRACTITIONERCOUNT + I(PRACTITIONERCOUNT^2)
  + I(PRACTITIONERCOUNT^3) +
  I(PRACTITIONERCOUNT^4) + recelig + PRACTITIONERCOUNT * recelig
  + I(PRACTITIONERCOUNT^2) *
  recelig + I(PRACTITIONERCOUNT^3) * recelig + I(
  PRACTITIONERCOUNT^4) * recelig,
  data = pract.sub[which(pract.sub$PRACTITIONERCOUNT ≤ 15 &
  pract.sub$PRACTITIONERCOUNT ≥
  5), ])

rec_quart10 ← lm(EHR ~ PRACTITIONERCOUNT + I(PRACTITIONERCOUNT^2)
  + I(PRACTITIONERCOUNT^3) +
  I(PRACTITIONERCOUNT^4) + recelig + PRACTITIONERCOUNT * recelig
  + I(PRACTITIONERCOUNT^2) *
  recelig + I(PRACTITIONERCOUNT^3) * recelig + I(
  PRACTITIONERCOUNT^4) * recelig,
  data = pract.sub[which(pract.sub$PRACTITIONERCOUNT ≤ 20), ])

# rec_disptbl_pt1← data.frame(x=seq(0,10,by=0.01),y1_5=0,y1_10=0,
  y2=0,y3=0,
# y4=0)
# rec_disptbl_pt2← data.frame(x=seq(10,20,by=0.01),y1_5=0,y1_10=0,
  y2=0,y3=0,
# y4=0)
coefs_rec_lin_bw5← coefficients(rec_lin_bw5)
# coefs_rec_lin_bw10← coefficients(rec_lin_bw10)
# coefs_rec_quad← coefficients(rec_quad)
# coefs_rec_cub← coefficients(rec_cub)
# coefs_rec_quart← coefficients(rec_quart) #calculate linear bw=5
  estimates
# rec_disptbl_pt1$y1_5← coefs_rec_lin_bw5[1]+coefs_rec_lin_bw5[2]*
  rec_disptbl_pt1$x+(coefs_rec_lin_bw5[2]+coefs_rec_lin_bw5[4])*

```

```

    rec_disptbl_pt1$x+coefs_rec_lin_bw5[3]
# rec_disptbl_pt2$y1_5← coefs_rec_lin_bw5[1]+coefs_rec_lin_bw5[2]*
  rec_disptbl_pt2$x
# #calculate linear bw=10 estimates
# rec_disptbl_pt1$y1_10← coefs_rec_lin_bw10[1]+(coefs_rec_lin_bw10
  [2]+coefs_rec_lin_bw10[4])*rec_disptbl_pt1$x+
  coefs_rec_lin_bw10[3]
# rec_disptbl_pt2$y1_10← coefs_rec_lin_bw10[1]+coefs_rec_lin_bw10
  [2]*rec_disptbl_pt2$x
# #calculate quadratic estimates
# rec_disptbl_pt1$y2← coefs_rec_quad[1]+(coefs_rec_quad[2]+
  coefs_rec_quad[5])*rec_disptbl_pt1$x+(coefs_rec_quad[3]+
  coefs_rec_quad[6])*I(rec_disptbl_pt1$x^2)+coefs_rec_quad[4]
# rec_disptbl_pt2$y2← coefs_rec_quad[1]+coefs_rec_quad[2]*
  rec_disptbl_pt2$x+coefs_rec_quad[3]*I(rec_disptbl_pt2$x^2)
# #calculate cubic estimates
# rec_disptbl_pt1$y3← coefs_rec_cub[1]+(coefs_rec_cub[2]+
  coefs_rec_cub[6])*rec_disptbl_pt1$x+(coefs_rec_cub[3]+
  coefs_rec_cub[7])*I(rec_disptbl_pt1$x^2)+(coefs_rec_cub[4]+
  coefs_rec_cub[8])*I(rec_disptbl_pt1$x^3)+coefs_rec_cub[5]
# rec_disptbl_pt2$y3← coefs_rec_cub[1]+coefs_rec_cub[2]*
  rec_disptbl_pt2$x+coefs_rec_cub[3]*I(rec_disptbl_pt2$x^2)+
  coefs_rec_cub[4]*I(rec_disptbl_pt2$x^3)
# #calculate quartic estimates
# rec_disptbl_pt1$y4← coefs_rec_quart[1]+(coefs_rec_quart[2]+
  coefs_rec_quart[7])*rec_disptbl_pt1$x+(coefs_rec_quart[3]+
  coefs_rec_quart[8])*I(rec_disptbl_pt1$x^2)+(coefs_rec_quart
  [4]+coefs_rec_quart[9])*I(rec_disptbl_pt1$x^3)+(
  coefs_rec_quart[5]+coefs_rec_quart[10])*I(rec_disptbl_pt1$x^4)
+coefs_rec_quart[6]
# rec_disptbl_pt2$y4← coefs_rec_quart[1]+coefs_rec_quart[2]*
  rec_disptbl_pt2$x+coefs_rec_quart[3]*I(rec_disptbl_pt2$x^2)+
  coefs_rec_quart[4]*I(rec_disptbl_pt2$x^3)+coefs_rec_quart[5]*I
  (rec_disptbl_pt2$x^4)

```

```

# *R Code Block* - recefftbl_bw

```

```

rec_eff_bw5 ← summary(rec_lin_bw5)$coefficients[3, c(1, 4)]
rec_eff_bw10 ← summary(rec_lin_bw10)$coefficients[3, c(1, 4)]
rec_eff_quad ← summary(rec_quad)$coefficients[4, c(1, 4)]
rec_eff_cub ← summary(rec_cub)$coefficients[5, c(1, 4)]
rec_eff_quart ← summary(rec_quart)$coefficients[6, c(1, 4)]

# rec_eff_bw1← summary(rec_lin_bw1)$coefficients[3, c(1, 4)]
rec_eff_bw3 ← summary(rec_lin_bw3)$coefficients[3, c(1, 4)]

```

```

rec_eff_quad5 ← summary(rec_quad5)$coefficients[4, c(1, 4)]
rec_eff_quad10 ← summary(rec_quad10)$coefficients[4, c(1, 4)]
rec_eff_cub5 ← summary(rec_cub5)$coefficients[5, c(1, 4)]
rec_eff_cub10 ← summary(rec_cub10)$coefficients[5, c(1, 4)]
rec_eff_quart5 ← summary(rec_quart5)$coefficients[6, c(1, 4)]
rec_eff_quart10 ← summary(rec_quart10)$coefficients[6, c(1, 4)]

rec_eff ← rbind(rec_eff_bw3, rec_eff_bw5, rec_eff_bw10,
  rec_eff_quad, rec_eff_quad5,
  rec_eff_quad10, rec_eff_cub, rec_eff_cub5, rec_eff_cub10,
  rec_eff_quart,
  rec_eff_quart5, rec_eff_quart10)
rownames(rec_eff) ← c("Local linear, bandwidth=3", "Local linear,
  bandwidth=5",
  "Local linear, bandwidth=10", "Quadratic", "Quadratic,
  bandwidth=5", "Quadratic, bandwidth=10",
  "Cubic", "Cubic, bandwidth=5", "Cubic, bandwidth=10", "Quartic
  ", "Quartic, bandwidth=5",
  "Quartic, bandwidth=10")

xtable(rec_eff, type = "latex", align = "|r|r|r|", asis = "TRUE",
  floating = FALSE,
  digits = 4, label = "rec_eff_exp", caption = "Expanded REC RDD
  Effect Sizes at Cut-off")

# ## ---- disprecpolyplot_bw # *R Code Block* - disprecpolyplot_bw
# ggplot()
# + geom_line(data=rec_disptbl_pt1, aes(x=x, y=y1_5, color='Linear,
#   bw=5')) +
# geom_line(data=rec_disptbl_pt2, aes(x=x, y=y1_5, color='Linear,
#   bw=5')) +
# geom_line(data=rec_disptbl_pt1, aes(x=x, y=y1_10, color='Linear,
#   bw=10')) +
# geom_line(data=rec_disptbl_pt2, aes(x=x, y=y1_10, color='Linear,
#   bw=10')) +
# geom_smooth(data=rec_disptbl_pt1, aes(x=x, y=y2, color='
#   Quadratic'),
# method='lm', formula=y~x+I(x^2)) + geom_smooth(data=
#   rec_disptbl_pt2,
# aes(x=x, y=y2, color='Quadratic'), method='lm', formula=y~x+I(x^
#   2)) +
# geom_smooth(data=rec_disptbl_pt1, aes(x=x, y=y3, color='Cubic'),
# method='lm', formula=y~x+I(x^2)+I(x^3)) +

```

```

# geom_smooth(data=rec_disptbl_pt2, aes(x=x, y=y3, color='Cubic'),
# method='lm', formula=y~x+I(x^2)+I(x^3)) +
# geom_smooth(data=rec_disptbl_pt1, aes(x=x, y=y4, color='Quartic
'),
# method='lm', formula=y~x+I(x^2)+I(x^3)+I(x^4)) +
# geom_smooth(data=rec_disptbl_pt2, aes(x=x, y=y4, color='Quartic
'),
# method='lm', formula=y~x+I(x^2)+I(x^3)+I(x^4)) + geom_vline(
  xintercept=10,
# linetype='dotted') + geom_rect(mapping=aes(xmin=8, xmax=12, ymin
  = 0,
# ymax=1.2), fill=NA, color='black', linetype='dotted') +
# scale_color_discrete(name='Model', breaks = c('Linear, bw=5', '
  Linear,
# bw=10', 'Quadratic', 'Cubic', 'Quartic')) ## ----
  disprecpolyplotzoom_bw #
# *R Code Block* - disprecpolyplotzoom_bw ggplot() +
# geom_line(data=rec_disptbl_pt1, aes(x=x, y=y1_5, color='Linear,
  bw=5')) +
# geom_line(data=rec_disptbl_pt2, aes(x=x, y=y1_5, color='Linear,
  bw=5')) +
# geom_line(data=rec_disptbl_pt1, aes(x=x, y=y1_10, color='Linear,
  bw=10'))
# + geom_line(data=rec_disptbl_pt2, aes(x=x, y=y1_10, color='
  Linear,
# bw=10')) + geom_smooth(data=rec_disptbl_pt1, aes(x=x, y=y2,
# color='Quadratic'), method='lm', formula=y~x+I(x^2)) +
# geom_smooth(data=rec_disptbl_pt2, aes(x=x, y=y2, color='
  Quadratic'),
# method='lm', formula=y~x+I(x^2)) + geom_smooth(data=
  rec_disptbl_pt1,
# aes(x=x, y=y3, color='Cubic'), method='lm', formula=y~x+I(x^2)+I
  (x^3)) +
# geom_smooth(data=rec_disptbl_pt2, aes(x=x, y=y3, color='Cubic'),
# method='lm', formula=y~x+I(x^2)+I(x^3)) +
# geom_smooth(data=rec_disptbl_pt1, aes(x=x, y=y4, color='Quartic
'),
# method='lm', formula=y~x+I(x^2)+I(x^3)+I(x^4)) +
# geom_smooth(data=rec_disptbl_pt2, aes(x=x, y=y4, color='Quartic
'),
# method='lm', formula=y~x+I(x^2)+I(x^3)+I(x^4)) + geom_vline(
  xintercept=10,
# linetype='dotted') + xlim(8,12) + ylim(0,1.2)+
# scale_color_discrete(name='Model', breaks = c('Linear, bw=5', '
  Linear,

```



```
# bw=10', 'Quadratic', 'Cubic', 'Quartic'))
```

## K.14 R Code Used for Statistics, Tables, and Figures in Appendix I - Robustness of Hospital RD Design Models to Addition of Covariates

```
# *R Code Block* - loadlibraries - Load non-base R libraries
rm(list = ls())
library(stringr)
library(xtable)
library(ggplot2)
library(plyr)
library(corrgram)
library(stargazer)
library(BaylorEdPsych)
library(sas7bdat)

# *R Code Block* - init
figtheme ← theme(plot.title = element_text(face = "bold", size =
  16)) + theme(axis.title.y = element_text(face = "bold",
  size = 14), axis.text.y = element_text(size = 12, color = "
  black"), axis.title.x = element_text(face = "bold",
  size = 14), axis.text.x = element_text(size = 12, color = "
  black"))

dir ← "C"

# *R Code Block* - makehospfile for hospital 2012 survey

hosp ← read.csv(file = str_c(dir, ":\users\scptex\Dropbox\
  Dissertation\Data\2012HospSurvey\HospSurv2012.csv"),
  head = TRUE, sep = ",")

hosp$EHR ← sub(2, 1, hosp$EHR)
hosp$EHR ← sub(3, 0, hosp$EHR)
hosp$EHR ← sub(4, 0, hosp$EHR)
hosp$EHR ← sub(5, 0, hosp$EHR)
hosp$EHR ← sub(6, 0, hosp$EHR)
hosp$EHR ← as.factor(hosp$EHR)

hosplink ← read.csv(file = str_c(dir, ":\users\scptex\Dropbox
  \Dissertation\Data\Hosp2012FromDSHS\HS12CD.csv"),
  head = TRUE, sep = ",")

hosplink2 ← read.csv(file = str_c(dir, ":\users\scptex\Dropbox
  \Dissertation\Data\Hosp2012FromDSHS\HS12DIR.csv"),
  head = TRUE, sep = ",")
```

```

hosp ← merge(hosp, hosplink, by = "FID")

hosp ← merge(hosp, hosplink2, by = "FID")

source(str_c(dir, ":\users\scptex\Dropbox\Dissertation\Text
  \code\counties\LoadCountyFiles.r"))

counties$Name ← toupper(counties$Name)
counties$Name ← gsub(" COUNTY", "", counties$Name)
counties$Name ← gsub(" ", "", counties$Name)
hosp$county ← gsub(" ", "", hosp$COUNTY)

hosp_comerge ← merge(hosp, counties, by.x = "COUNTY", by.y = "
  Name")

hosp_comerge$MEDIANINCOME ← as.numeric(gsub(",", "", as.character
  (hosp_comerge$MEDIANINCOME)))
hosp_comerge$POVERTYPCT ← as.numeric(gsub(",", "", as.character(
  hosp_comerge$POVERTYPCT)))

HRRZipXref ← read.csv(file = str_c(dir, ":\users\scptex\
  Dropbox\Dissertation\Data\ DartmouthHRRs\ZipHsaHrr11.csv"
  ),
  head = TRUE, sep = ",") #read in HRR-zip xref file

hosp_hrrmerge ← merge(hosp_comerge, HRRZipXref, by.x = "LOCOZIP",
  by.y = "zipcode11")

hosp_sub ← hosp_hrrmerge[, c("FID", "FacName", "EHR", "EHYear", "
  D1A1", "D1F1",
  "D1G1", "D2A11", "D2B11", "D2C11", "D2D11", "D3E1", "D5B1", "
  MSTAT", "POPDENSITY",
  "MEDIANINCOME", "POVERTYPCT", "hrrnum")]

names(hosp_sub) ← c("fid", "facname", "ehr", "ehryear", "
  totalbeds", "inpatientdays",
  "edvisits", "medicaredischarges", "medicaredays", "
  medicaiddischarges",
  "medicaiddays", "totalrevenue", "charitycare", "metrostat", "
  popdensity",
  "medianincome", "povertypct", "hrrnum")

hosp_sub$size ← cut(hosp_sub$totalbeds, labels = c("Small", "
  Medium", "Large"),

```

```

    breaks = c(0, 100, 400, 5000))

hosp_sub$metrostat ← gsub("1", "metro", hosp_sub$metrostat)
hosp_sub$metrostat ← gsub("2", "nonmetro", hosp_sub$metrostat)
hosp_sub$povpct_lvls ← cut(hosp_sub$povertypct, 4)
hosp_sub$metrostat ← factor(hosp_sub$metrostat)

hosp_sub$mcdpct ← hosp_sub$medicaiddays/hosp_sub$inpatientdays
hosp_sub$mcdehrelig ← ifelse(hosp_sub$mcdpct > 0.1, 1, 0)

hosp_sub1 ← hosp_sub[hosp_sub$ehr == 0, ]
hosp_sub2 ← hosp_sub[hosp_sub$ehr == 1 & hosp_sub$ehryear ≥
  2009, ]
hosp_sub ← rbind(hosp_sub1, hosp_sub2)

# *R Code Block* - calcmcdpolyplots
hosp_sub$ehr ← as.numeric(hosp_sub$ehr) - 1

# Imbens-Kalyanaraman Optimal Bandwidth Calculation
mcdIKbw ← IKbandwidth(hosp_sub$mcdpct, hosp_sub$ehr, cutpoint = 0
  .1)

# DCdensity(hosp_sub$mcdpct, cutpoint=0.1, bin=0.01)

mcd_lin_bw5 ← lm(ehr ~ mcdpct + mcdehrelig + metrostat +
  povpct_lvls + mcdpct *
  mcdehrelig, data = hosp_sub[which(hosp_sub$mcdpct < 0.15 &
  hosp_sub$mcdpct >
  0.05), ])

mcd_lin_bw10 ← lm(ehr ~ mcdpct + mcdehrelig + metrostat +
  povpct_lvls + mcdpct *
  mcdehrelig, data = hosp_sub[which(hosp_sub$mcdpct < 0.2), ])

mcd_quad ← lm(ehr ~ mcdpct + I(mcdpct^2) + mcdehrelig + metrostat
  + povpct_lvls +
  mcdpct * mcdehrelig + I(mcdpct^2) * mcdehrelig, data =
  hosp_sub[which(hosp_sub$mcdpct <
  0.4), ])

mcd_cub ← lm(ehr ~ mcdpct + I(mcdpct^2) + I(mcdpct^3) +
  mcdehrelig + metrostat +
  povpct_lvls + mcdpct * mcdehrelig + I(mcdpct^2) * mcdehrelig +
  I(mcdpct^3) *
  mcdehrelig, data = hosp_sub[which(hosp_sub$mcdpct < 0.4), ])

```

```

mcd_quart <- lm(ehr ~ mcdpct + I(mcdpct^2) + I(mcdpct^3) + I(
  mcdpct^4) + mcdehrelig +
  metrostat + povpct_lvls + mcdpct * mcdehrelig + I(mcdpct^2) *
  mcdehrelig +
  I(mcdpct^3) * mcdehrelig + I(mcdpct^4) * mcdehrelig, data =
  hosp_sub[which(hosp_sub$mcdpct <
  0.4), ])

mcd_disptbl_pt1 <- data.frame(x = seq(0, 0.1, by = 0.001), y1_5 =
  0, y1_10 = 0,
  y2 = 0, y3 = 0, y4 = 0)
mcd_disptbl_pt2 <- data.frame(x = seq(0.1, 0.2, by = 0.001), y1_5
  = 0, y1_10 = 0,
  y2 = 0, y3 = 0, y4 = 0)

coefs_mcd_lin_bw5 <- coefficients(mcd_lin_bw5)
coefs_mcd_lin_bw10 <- coefficients(mcd_lin_bw10)
coefs_mcd_quad <- coefficients(mcd_quad)
coefs_mcd_cub <- coefficients(mcd_cub)
coefs_mcd_quart <- coefficients(mcd_quart)

# calculate linear bw=5 estimates
mcd_disptbl_pt1$y1_5 <- coefs_mcd_lin_bw5[1] + coefs_mcd_lin_bw5
  [2] * mcd_disptbl_pt1$x
mcd_disptbl_pt2$y1_5 <- coefs_mcd_lin_bw5[1] + coefs_mcd_lin_bw5
  [2] * mcd_disptbl_pt2$x +
  coefs_mcd_lin_bw5[3]

# calculate linear bw=10 estimates
mcd_disptbl_pt1$y1_10 <- coefs_mcd_lin_bw10[1] +
  coefs_mcd_lin_bw10[2] * mcd_disptbl_pt1$x
mcd_disptbl_pt2$y1_10 <- coefs_mcd_lin_bw10[1] +
  coefs_mcd_lin_bw10[2] * mcd_disptbl_pt2$x +
  coefs_mcd_lin_bw10[3]

# calculate quadratic estimates
mcd_disptbl_pt1$y2 <- coefs_mcd_quad[1] + coefs_mcd_quad[2] *
  mcd_disptbl_pt1$x +
  coefs_mcd_quad[3] * I(mcd_disptbl_pt1$x^2)
mcd_disptbl_pt2$y2 <- coefs_mcd_quad[1] + (coefs_mcd_quad[2] +
  coefs_mcd_quad[5]) *
  mcd_disptbl_pt2$x + (coefs_mcd_quad[3] + coefs_mcd_quad[6]) *
  I(mcd_disptbl_pt2$x^2) +
  coefs_mcd_quad[4]

```

```

# calculate cubic estimates
mcd_disptbl_pt1$y3 ← coefs_mcd_cub[1] + coefs_mcd_cub[2] *
  mcd_disptbl_pt1$x +
  coefs_mcd_cub[3] * I(mcd_disptbl_pt1$x^2) + coefs_mcd_cub[4] *
  I(mcd_disptbl_pt1$x^3)
mcd_disptbl_pt2$y3 ← coefs_mcd_cub[1] + (coefs_mcd_cub[2] +
  coefs_mcd_cub[6]) *
  mcd_disptbl_pt2$x + (coefs_mcd_cub[3] + coefs_mcd_cub[7]) * I(
  mcd_disptbl_pt2$x^2) +
  (coefs_mcd_cub[4] + coefs_mcd_cub[8]) * I(mcd_disptbl_pt2$x^3)
  + coefs_mcd_cub[5]

# calculate quartic estimates
mcd_disptbl_pt1$y4 ← coefs_mcd_quart[1] + coefs_mcd_quart[2] *
  mcd_disptbl_pt1$x +
  coefs_mcd_quart[3] * I(mcd_disptbl_pt1$x^2) + coefs_mcd_quart
  [4] * I(mcd_disptbl_pt1$x^3) +
  coefs_mcd_quart[5] * I(mcd_disptbl_pt1$x^4)
mcd_disptbl_pt2$y4 ← coefs_mcd_quart[1] + (coefs_mcd_quart[2] +
  coefs_mcd_quart[7]) *
  mcd_disptbl_pt2$x + (coefs_mcd_quart[3] + coefs_mcd_quart[8])
  * I(mcd_disptbl_pt2$x^2) +
  (coefs_mcd_quart[4] + coefs_mcd_quart[9]) * I(mcd_disptbl_pt2$
  x^3) + (coefs_mcd_quart[5] +
  coefs_mcd_quart[10]) * I(mcd_disptbl_pt2$x^4) +
  coefs_mcd_quart[6]

# *R Code Block* - dispmcdpolyplot
ggplot() + geom_line(data = mcd_disptbl_pt1, aes(x = x, y = y1.5,
  color = "Linear", bw=5)) +
  geom_line(data = mcd_disptbl_pt2, aes(x = x, y = y1.5, color =
  "Linear", bw=5)) +
  geom_line(data = mcd_disptbl_pt1, aes(x = x, y = y1.10, color
  = "Linear", bw=10)) +
  geom_line(data = mcd_disptbl_pt2, aes(x = x, y = y1.10, color
  = "Linear", bw=10)) +
  geom_smooth(data = mcd_disptbl_pt1, aes(x = x, y = y2, color =
  "Quadratic"),
  method = "lm", formula = y ~ x + I(x^2)) + geom_smooth(
  data = mcd_disptbl_pt2,
  aes(x = x, y = y2, color = "Quadratic"), method = "lm",
  formula = y ~ x +
  I(x^2)) + geom_smooth(data = mcd_disptbl_pt1, aes(x = x, y
  = y3, color = "Cubic"),

```

```

method = "lm", formula = y ~ x + I(x^2) + I(x^3)) +
  geom_smooth(data = mcd_disptbl_pt2 ,
aes(x = x, y = y3, color = "Cubic"), method = "lm", formula =
  y ~ x + I(x^2) +
  I(x^3)) + geom_smooth(data = mcd_disptbl_pt1 , aes(x = x, y
  = y4, color = "Quartic"),
method = "lm", formula = y ~ x + I(x^2) + I(x^3) + I(x^4)) +
  geom_smooth(data = mcd_disptbl_pt2 ,
aes(x = x, y = y4, color = "Quartic"), method = "lm", formula
  = y ~ x +
  I(x^2) + I(x^3) + I(x^4)) + geom_vline(xintercept = 0.1 ,
  linetype = "dotted") +
scale_color_discrete(name = "Model", breaks = c("Linear", bw=0
  .05", "Linear", bw=0.10",
  "Quadratic", "Cubic", "Quartic")) + geom_rect(mapping =
  aes(xmin = 0.075 ,
xmax = 0.125, ymin = 0, ymax = 1), fill = NA, color = "black",
  linetype = "dotted") +
ylim(-0.25, 1.75)

```

```

# *R Code Block* - dispmcdpolyplotzoom
ggplot() + geom_line(data = mcd_disptbl_pt1 , aes(x = x, y = y1_5 ,
  color = "Linear", bw=5")) +
  geom_line(data = mcd_disptbl_pt2 , aes(x = x, y = y1_5 , color =
  "Linear", bw=5")) +
  geom_line(data = mcd_disptbl_pt1 , aes(x = x, y = y1_10 , color
  = "Linear", bw=10")) +
  geom_line(data = mcd_disptbl_pt2 , aes(x = x, y = y1_10 , color
  = "Linear", bw=10")) +
  geom_smooth(data = mcd_disptbl_pt1 , aes(x = x, y = y2, color =
  "Quadratic"),
  method = "lm", formula = y ~ x + I(x^2)) + geom_smooth(
  data = mcd_disptbl_pt2 ,
aes(x = x, y = y2, color = "Quadratic"), method = "lm",
  formula = y ~ x +
  I(x^2)) + geom_smooth(data = mcd_disptbl_pt1 , aes(x = x, y
  = y3, color = "Cubic"),
method = "lm", formula = y ~ x + I(x^2) + I(x^3)) +
  geom_smooth(data = mcd_disptbl_pt2 ,
aes(x = x, y = y3, color = "Cubic"), method = "lm", formula =
  y ~ x + I(x^2) +
  I(x^3)) + geom_smooth(data = mcd_disptbl_pt1 , aes(x = x, y
  = y4, color = "Quartic"),
method = "lm", formula = y ~ x + I(x^2) + I(x^3) + I(x^4)) +
  geom_smooth(data = mcd_disptbl_pt2 ,

```

```

aes(x = x, y = y4, color = "Quartic"), method = "lm", formula
= y ~ x +
  I(x^2) + I(x^3) + I(x^4)) + geom_vline(xintercept = 0.1,
  linetype = "dotted") +
scale_color_discrete(name = "Model", breaks = c("Linear", bw=5"
, "Linear", bw=10",
"Quadratic", "Cubic", "Quartic")) + xlim(0.075, 0.125) +
ylim(0, 1)

```

```

# *R Code Block* - mcdefftbl
mcd_eff_bw5 <- summary(mcd_lin_bw5)$coefficients[3, c(1, 4)]
mcd_eff_bw10 <- summary(mcd_lin_bw10)$coefficients[3, c(1, 4)]
mcd_eff_quad <- summary(mcd_quad)$coefficients[4, c(1, 4)]
mcd_eff_cub <- summary(mcd_cub)$coefficients[5, c(1, 4)]
mcd_eff_quart <- summary(mcd_quart)$coefficients[6, c(1, 4)]

mcd_eff <- rbind(mcd_eff_bw5, mcd_eff_bw10, mcd_eff_quad,
  mcd_eff_cub, mcd_eff_quart)
rownames(mcd_eff) <- c("Local linear, bandwidth=5", "Local linear,
  bandwidth=10",
  "Quadratic", "Cubic", "Quartic")

xtable(mcd_eff, type = "latex", align = "|r|r|r|", asis = "TRUE",
  floating = FALSE,
  digits = 4, label = "mcd_eff", caption = "Medicaid RDD Effect
  Sizes at Cut-off with Covariates")

```



## K.15 R Code Used for Statistics, Tables, and Figures in Appendix J - Robustness of Hospital RD Design Models to Different Bandwidth Calculations

```
# *R Code Block* - loadlibraries - Load non-base R libraries
rm(list = ls())
library(stringr)
library(xtable)
library(ggplot2)
library(plyr)
library(corrgram)
library(stargazer)
library(BaylorEdPsych)
library(sas7bdat)

# *R Code Block* - init
figtheme ← theme(plot.title = element_text(face = "bold", size =
  16)) + theme(axis.title.y = element_text(face = "bold",
  size = 14), axis.text.y = element_text(size = 12, color = "
  black"), axis.title.x = element_text(face = "bold",
  size = 14), axis.text.x = element_text(size = 12, color = "
  black"))

dir ← "C"

# *R Code Block* - makehospfile for hospital 2012 survey

hosp ← read.csv(file = str_c(dir, ":\users\scptex\Dropbox\
  Dissertation\Data\2012HospSurvey\HospSurv2012.csv"),
  head = TRUE, sep = ",")

hosp$EHR ← sub(2, 1, hosp$EHR)
hosp$EHR ← sub(3, 0, hosp$EHR)
hosp$EHR ← sub(4, 0, hosp$EHR)
hosp$EHR ← sub(5, 0, hosp$EHR)
hosp$EHR ← sub(6, 0, hosp$EHR)
hosp$EHR ← as.factor(hosp$EHR)

hosplink ← read.csv(file = str_c(dir, ":\users\scptex\Dropbox
  \Dissertation\Data\Hosp2012FromDSHS\HS12CD.csv"),
  head = TRUE, sep = ",")

hosplink2 ← read.csv(file = str_c(dir, ":\users\scptex\Dropbox
  \Dissertation\Data\Hosp2012FromDSHS\HS12DIR.csv"),
  head = TRUE, sep = ",")
```

```

hosp ← merge(hosp, hosplink, by = "FID")

hosp ← merge(hosp, hosplink2, by = "FID")

source(str_c(dir, ":\users\scptex\Dropbox\Dissertation\Text
  \code\counties\LoadCountyFiles.r"))

counties$Name ← toupper(counties$Name)
counties$Name ← gsub(" COUNTY", "", counties$Name)
counties$Name ← gsub(" ", "", counties$Name)
hosp$county ← gsub(" ", "", hosp$COUNTY)

hosp_comerge ← merge(hosp, counties, by.x = "COUNTY", by.y = "
  Name")

hosp_comerge$MEDIANINCOME ← as.numeric(gsub(",", "", as.character(
  hosp_comerge$MEDIANINCOME)))
hosp_comerge$POVERTYPCT ← as.numeric(gsub(",", "", as.character(
  hosp_comerge$POVERTYPCT)))

HRRZipXref ← read.csv(file = str_c(dir, ":\users\scptex\
  Dropbox\Dissertation\Data\ DartmouthHRRs\ZipHsaHrr11.csv"
  ),
  head = TRUE, sep = ",") #read in HRR-zip xref file

hosp_hrrmerge ← merge(hosp_comerge, HRRZipXref, by.x = "LOCOZIP",
  by.y = "zipcode11")

hosp_sub ← hosp_hrrmerge[, c("FID", "FacName", "EHR", "EHYear", "
  D1A1", "D1F1",
  "D1G1", "D2A11", "D2B11", "D2C11", "D2D11", "D3E1", "D5B1", "
  MSTAT", "POPDENSITY",
  "MEDIANINCOME", "POVERTYPCT", "hrrnum")]

names(hosp_sub) ← c("fid", "facname", "ehr", "ehryear", "
  totalbeds", "inpatientdays",
  "edvisits", "medicaredischarges", "medicaredays", "
  medicaiddischarges",
  "medicaiddays", "totalrevenue", "charitycare", "metrostat", "
  popdensity",
  "medianincome", "povertypct", "hrrnum")

hosp_sub$size ← cut(hosp_sub$totalbeds, labels = c("Small", "
  Medium", "Large"),

```

```

breaks = c(0, 100, 400, 5000))

hosp_sub$metrostat ← gsub("1", "metro", hosp_sub$metrostat)
hosp_sub$metrostat ← gsub("2", "nonmetro", hosp_sub$metrostat)
hosp_sub$povpct_lvls ← cut(hosp_sub$povertypct, 4)
hosp_sub$metrostat ← factor(hosp_sub$metrostat)

hosp_sub$mcdpct ← hosp_sub$medicaiddays/hosp_sub$inpatientdays
hosp_sub$mcdehrelig ← ifelse(hosp_sub$mcdpct > 0.1, 1, 0)

hosp_sub1 ← hosp_sub[hosp_sub$ehr == 0, ]
hosp_sub2 ← hosp_sub[hosp_sub$ehr == 1 & hosp_sub$ehryear ≥
  2009, ]
hosp_sub ← rbind(hosp_sub1, hosp_sub2)

# *R Code Block* - calcmcdpolyplots
hosp_sub$ehr ← as.numeric(hosp_sub$ehr) - 1

# Imbens-Kalyanaraman Optimal Bandwidth Calculation
mcdIKbw ← IKbandwidth(hosp_sub$mcdpct, hosp_sub$ehr, cutpoint = 0
  .1)

# DCdensity(hosp_sub$mcdpct, cutpoint=0.1, bin=0.01)

mcd_lin_bw1 ← lm(ehr ~ mcdpct + mcdehrelig + mcdpct * mcdehrelig,
  data = hosp_sub[which(hosp_sub$mcdpct ≤
  0.11 & hosp_sub$mcdpct ≥ 0.09), ])

mcd_lin_bw3 ← lm(ehr ~ mcdpct + mcdehrelig + mcdpct * mcdehrelig,
  data = hosp_sub[which(hosp_sub$mcdpct ≤
  0.13 & hosp_sub$mcdpct ≥ 0.07), ])

mcd_lin_bw5 ← lm(ehr ~ mcdpct + mcdehrelig + mcdpct * mcdehrelig,
  data = hosp_sub[which(hosp_sub$mcdpct <
  0.15 & hosp_sub$mcdpct > 0.05), ])

mcd_lin_bw10 ← lm(ehr ~ mcdpct + mcdehrelig + mcdpct * mcdehrelig
  , data = hosp_sub[which(hosp_sub$mcdpct <
  0.2), ])

mcd_quad ← lm(ehr ~ mcdpct + I(mcdpct^2) + mcdehrelig + mcdpct *
  mcdehrelig +
  I(mcdpct^2) * mcdehrelig, data = hosp_sub[which(hosp_sub$
  mcdpct < 0.4),
  ])

```

```

mcd_quad5 ← lm(ehr ~ mcdpct + I(mcdpct^2) + mcdehrelig + mcdpct *
  mcdehrelig +
  I(mcdpct^2) * mcdehrelig, data = hosp_sub[which(hosp_sub$
    mcdpct < 0.15 &
    hosp_sub$mcdpct > 0.05), ])

mcd_quad10 ← lm(ehr ~ mcdpct + I(mcdpct^2) + mcdehrelig + mcdpct
  * mcdehrelig +
  I(mcdpct^2) * mcdehrelig, data = hosp_sub[which(hosp_sub$
    mcdpct < 0.2),
  ])

mcd_cub ← lm(ehr ~ mcdpct + I(mcdpct^2) + I(mcdpct^3) +
  mcdehrelig + mcdpct *
  mcdehrelig + I(mcdpct^2) * mcdehrelig + I(mcdpct^3) *
  mcdehrelig, data = hosp_sub[which(hosp_sub$mcdpct <
    0.4), ])

mcd_cub5 ← lm(ehr ~ mcdpct + I(mcdpct^2) + I(mcdpct^3) +
  mcdehrelig + mcdpct *
  mcdehrelig + I(mcdpct^2) * mcdehrelig + I(mcdpct^3) *
  mcdehrelig, data = hosp_sub[which(hosp_sub$mcdpct <
    0.15 & hosp_sub$mcdpct > 0.05), ])

mcd_cub10 ← lm(ehr ~ mcdpct + I(mcdpct^2) + I(mcdpct^3) +
  mcdehrelig + mcdpct *
  mcdehrelig + I(mcdpct^2) * mcdehrelig + I(mcdpct^3) *
  mcdehrelig, data = hosp_sub[which(hosp_sub$mcdpct <
    0.2), ])

mcd_quart ← lm(ehr ~ mcdpct + I(mcdpct^2) + I(mcdpct^3) + I(
  mcdpct^4) + mcdehrelig +
  mcdpct * mcdehrelig + I(mcdpct^2) * mcdehrelig + I(mcdpct^3) *
  mcdehrelig +
  I(mcdpct^4) * mcdehrelig, data = hosp_sub[which(hosp_sub$
    mcdpct < 0.4),
  ])

mcd_quart5 ← lm(ehr ~ mcdpct + I(mcdpct^2) + I(mcdpct^3) + I(
  mcdpct^4) + mcdehrelig +
  mcdpct * mcdehrelig + I(mcdpct^2) * mcdehrelig + I(mcdpct^3) *
  mcdehrelig +
  I(mcdpct^4) * mcdehrelig, data = hosp_sub[which(hosp_sub$
    mcdpct < 0.15 &

```

```

hosp_sub$mcdpct > 0.05), ])

mcd_quart10 ← lm(ehr ~ mcdpct + I(mcdpct^2) + I(mcdpct^3) + I(
  mcdpct^4) + mcdehrelig +
  mcdpct * mcdehrelig + I(mcdpct^2) * mcdehrelig + I(mcdpct^3) *
  mcdehrelig +
  I(mcdpct^4) * mcdehrelig, data = hosp_sub[which(hosp_sub$
  mcdpct < 0.2),
  ])

# mcd_disptbl_pt1 ← data.frame(x=seq(0,0.10,by=0.001),y1_5=0,y1_10
  =0,y2=0,y3=0,
# y4=0)
# mcd_disptbl_pt2 ← data.frame(x=seq(0.10,0.20,by=0.001),y1_5=0,
  y1_10=0,y2=0,y3=0,
# y4=0)
coefs_mcd_lin_bw5 ← coefficients(mcd_lin_bw5)
# coefs_mcd_lin_bw10 ← coefficients(mcd_lin_bw10)
# coefs_mcd_quad ← coefficients(mcd_quad)
# coefs_mcd_cub ← coefficients(mcd_cub)
# coefs_mcd_quart ← coefficients(mcd_quart) #calculate linear bw=5
  estimates
# mcd_disptbl_pt1$y1_5 ← coefs_mcd_lin_bw5[1]+coefs_mcd_lin_bw5[2]*
  mcd_disptbl_pt1$x
# mcd_disptbl_pt2$y1_5 ← coefs_mcd_lin_bw5[1]+coefs_mcd_lin_bw5[2]*
  mcd_disptbl_pt2$x+coefs_mcd_lin_bw5[3]
# #calculate linear bw=10 estimates
# mcd_disptbl_pt1$y1_10 ← coefs_mcd_lin_bw10[1]+coefs_mcd_lin_bw10
  [2]*mcd_disptbl_pt1$x
# mcd_disptbl_pt2$y1_10 ← coefs_mcd_lin_bw10[1]+coefs_mcd_lin_bw10
  [2]*mcd_disptbl_pt2$x+coefs_mcd_lin_bw10[3]
# #calculate quadratic estimates
# mcd_disptbl_pt1$y2 ← coefs_mcd_quad[1]+coefs_mcd_quad[2]*
  mcd_disptbl_pt1$x+coefs_mcd_quad[3]*I(mcd_disptbl_pt1$x^2)
# mcd_disptbl_pt2$y2 ← coefs_mcd_quad[1]+(coefs_mcd_quad[2]+
  coefs_mcd_quad[5])*mcd_disptbl_pt2$x+(coefs_mcd_quad[3]+
  coefs_mcd_quad[6])*I(mcd_disptbl_pt2$x^2)+coefs_mcd_quad[4]
# #calculate cubic estimates
# mcd_disptbl_pt1$y3 ← coefs_mcd_cub[1]+coefs_mcd_cub[2]*
  mcd_disptbl_pt1$x+coefs_mcd_cub[3]*I(mcd_disptbl_pt1$x^2)+
  coefs_mcd_cub[4]*I(mcd_disptbl_pt1$x^3)
# mcd_disptbl_pt2$y3 ← coefs_mcd_cub[1]+(coefs_mcd_cub[2]+
  coefs_mcd_cub[6])*mcd_disptbl_pt2$x+(coefs_mcd_cub[3]+
  coefs_mcd_cub[7])*I(mcd_disptbl_pt2$x^2)+(coefs_mcd_cub[4]+
  coefs_mcd_cub[8])*I(mcd_disptbl_pt2$x^3)+coefs_mcd_cub[5]
# #calculate quartic estimates

```

```

# mcd_disptbl_pt1$y4 ← coefs_mcd_quart [1]+coefs_mcd_quart [2]*
  mcd_disptbl_pt1$x+coefs_mcd_quart [3]*I(mcd_disptbl_pt1$x^2)+
  coefs_mcd_quart [4]*I(mcd_disptbl_pt1$x^3)+coefs_mcd_quart [5]*I
  (mcd_disptbl_pt1$x^4)
# mcd_disptbl_pt2$y4 ← coefs_mcd_quart [1]+(coefs_mcd_quart [2]+
  coefs_mcd_quart [7])*mcd_disptbl_pt2$x+(coefs_mcd_quart [3]+
  coefs_mcd_quart [8])*I(mcd_disptbl_pt2$x^2)+(coefs_mcd_quart
  [4]+coefs_mcd_quart [9])*I(mcd_disptbl_pt2$x^3)+(
  coefs_mcd_quart [5]+coefs_mcd_quart [10])*I(mcd_disptbl_pt2$x^4)
+coefs_mcd_quart [6]

# *R Code Block* - mcdefftbl
mcd_eff_bw5 ← summary(mcd_lin_bw5)$coefficients [3, c(1, 4)]
mcd_eff_bw10 ← summary(mcd_lin_bw10)$coefficients [3, c(1, 4)]
mcd_eff_quad ← summary(mcd_quad)$coefficients [4, c(1, 4)]
mcd_eff_cub ← summary(mcd_cub)$coefficients [5, c(1, 4)]
mcd_eff_quart ← summary(mcd_quart)$coefficients [6, c(1, 4)]

mcd_eff_bw1 ← summary(mcd_lin_bw1)$coefficients [3, c(1, 4)]
mcd_eff_bw3 ← summary(mcd_lin_bw3)$coefficients [3, c(1, 4)]
mcd_eff_quad5 ← summary(mcd_quad5)$coefficients [4, c(1, 4)]
mcd_eff_cub5 ← summary(mcd_cub5)$coefficients [5, c(1, 4)]
mcd_eff_quart5 ← summary(mcd_quart5)$coefficients [6, c(1, 4)]

mcd_eff_quad10 ← summary(mcd_quad10)$coefficients [4, c(1, 4)]
mcd_eff_cub10 ← summary(mcd_cub10)$coefficients [5, c(1, 4)]
mcd_eff_quart10 ← summary(mcd_quart10)$coefficients [6, c(1, 4)]

mcd_eff ← rbind(mcd_eff_bw1, mcd_eff_bw3, mcd_eff_bw5,
  mcd_eff_bw10, mcd_eff_quad,
  mcd_eff_quad5, mcd_eff_quad10, mcd_eff_cub, mcd_eff_cub5,
  mcd_eff_cub10,
  mcd_eff_quart, mcd_eff_quart5, mcd_eff_quart10)
rownames(mcd_eff) ← c("Local linear, bandwidth=1", "Local linear,
  bandwidth=3",
  "Local linear, bandwidth=5", "Local linear, bandwidth=10", "
  Quadratic",
  "Quadratic, bandwidth=5", "Quadratic, bandwidth=10", "Cubic",
  "Cubic, bandwidth=5",
  "Cubic, bandwidth=10", "Quartic", "Quartic, bandwidth=5", "
  Quartic, bandwidth=10")

xtable(mcd_eff, type = "latex", align = "|r|r|r|", asis = "TRUE",
  floating = FALSE,
  digits = 4, label = "mcd_eff", caption = "Medicaid RDD Effect

```

## Sizes at Cut-off")

```

## ---- dispmcdpolyplot_bw # *R Code Block* - dispmcdpolyplot
ggplot() +
# geom_line(data=mcd_disptbl_pt1, aes(x=x, y=y1_5, color='Linear',
  bw=5')) +
# geom_line(data=mcd_disptbl_pt2, aes(x=x, y=y1_5, color='Linear',
  bw=5')) +
# geom_line(data=mcd_disptbl_pt1, aes(x=x, y=y1_10, color='Linear',
  bw=10')) +
# geom_line(data=mcd_disptbl_pt2, aes(x=x, y=y1_10, color='Linear',
  bw=10')) +
# geom_smooth(data=mcd_disptbl_pt1, aes(x=x, y=y2, color='
  Quadratic'),
# method='lm', formula=y~x+I(x^2)) + geom_smooth(data=
  mcd_disptbl_pt2,
# aes(x=x, y=y2, color='Quadratic'), method='lm', formula=y~x+I(x^
  2)) +
# geom_smooth(data=mcd_disptbl_pt1, aes(x=x, y=y3, color='Cubic'),
# method='lm', formula=y~x+I(x^2)+I(x^3)) +
# geom_smooth(data=mcd_disptbl_pt2, aes(x=x, y=y3, color='Cubic'),
# method='lm', formula=y~x+I(x^2)+I(x^3)) +
# geom_smooth(data=mcd_disptbl_pt1, aes(x=x, y=y4, color='Quartic
  '),
# method='lm', formula=y~x+I(x^2)+I(x^3)+I(x^4)) +
# geom_smooth(data=mcd_disptbl_pt2, aes(x=x, y=y4, color='Quartic
  '),
# method='lm', formula=y~x+I(x^2)+I(x^3)+I(x^4)) +
# geom_vline(xintercept=0.10, linetype='dotted') +
# scale_color_discrete(name='Model', breaks = c('Linear, bw=0.05',
  'Linear,
# bw=0.10', 'Quadratic', 'Cubic', 'Quartic')) +
# geom_rect(mapping=aes(xmin=0.075, xmax=0.125, ymin = 0, ymax=1),
  fill=NA,
# color='black', linetype='dotted')+ ylim(-0.25,1.75) ## ----
# dispmcdpolyplotzoom_bw # *R Code Block* - dispmcdpolyplotzoom
ggplot() +
# geom_line(data=mcd_disptbl_pt1, aes(x=x, y=y1_5, color='Linear',
  bw=5')) +
# geom_line(data=mcd_disptbl_pt2, aes(x=x, y=y1_5, color='Linear',
  bw=5')) +
# geom_line(data=mcd_disptbl_pt1, aes(x=x, y=y1_10, color='Linear',
  bw=10')) +
# geom_line(data=mcd_disptbl_pt2, aes(x=x, y=y1_10, color='Linear,

```

```

    bw=10')) +
# geom_smooth(data=mcd_disptbl_pt1, aes(x=x, y=y2, color='
  Quadratic'),
# method='lm', formula=y~x+I(x^2)) + geom_smooth(data=
  mcd_disptbl_pt2,
# aes(x=x, y=y2, color='Quadratic'), method='lm', formula=y~x+I(x^
  2)) +
# geom_smooth(data=mcd_disptbl_pt1, aes(x=x, y=y3, color='Cubic'),
# method='lm', formula=y~x+I(x^2)+I(x^3)) +
# geom_smooth(data=mcd_disptbl_pt2, aes(x=x, y=y3, color='Cubic'),
# method='lm', formula=y~x+I(x^2)+I(x^3)) +
# geom_smooth(data=mcd_disptbl_pt1, aes(x=x, y=y4, color='Quartic
  '),
# method='lm', formula=y~x+I(x^2)+I(x^3)+I(x^4)) +
# geom_smooth(data=mcd_disptbl_pt2, aes(x=x, y=y4, color='Quartic
  '),
# method='lm', formula=y~x+I(x^2)+I(x^3)+I(x^4)) +
# geom_vline(xintercept=0.10, linetype='dotted') +
# scale_color_discrete(name='Model', breaks = c('Linear, bw=5', '
  Linear,
# bw=10', 'Quadratic', 'Cubic', 'Quartic')) + xlim(0.075, 0.125) +
  ylim(0,1)

```



## Bibliography

- Abdolrasulnia, Maziar et al. (2008). “Market effects on electronic health record adoption by physicians.” eng. In: *Health care management review* 33.3, pp. 243–252. ISSN: 1550-5030. DOI: 10.1097/01.HMR.0000324904.19272.c2. URL: <http://www.ncbi.nlm.nih.gov/pubmed/18580304>.
- Adler-Milstein, Julia et al. (2014). “More than half of US hospitals have at least a basic EHR, but stage 2 criteria remain challenging for most.” eng. In: *Health affairs (Project Hope)* 33.9, pp. 1664–1671. ISSN: 1544-5208. DOI: 10.1377/hlthaff.2014.0453. URL: <http://www.ncbi.nlm.nih.gov/pubmed/25104826>.
- Andrews, James E. et al. (2004). “Current state of information technology use in a US primary care practice-based research network.” eng. In: *Informatics in primary care* 12.1, pp. 11–18. ISSN: 1476-0320. URL: <http://www.ncbi.nlm.nih.gov/pubmed/15140348>.
- Bass, Frank M. (1969). “A New Product Growth for Model Consumer Durables”. In: *Management Science* 15.5, pp. 215–227. URL: <http://EconPapers.repec.org/RePEc:inm:ormnsc:v:15:y:1969:i:5:p:215-227>.
- Bates, David W. (2005). “Physicians and ambulatory electronic health records.” eng. In: *Health affairs (Project Hope)* 24.5, pp. 1180–1189. ISSN: 0278-2715. DOI: 10.1377/hlthaff.24.5.1180. URL: <http://www.ncbi.nlm.nih.gov/pubmed/16162561>.
- Bazemore, Andrew et al. (2011). “Establishing a baseline: health information technology adoption among family medicine diplomates.” eng. In: *Journal of the American Board of Family Medicine : JABFM* 24.2, p. 132. ISSN: 1557-2625. DOI: 10.3122/jabfm.2011.02.110010. URL: <http://www.ncbi.nlm.nih.gov/pubmed/21383209>.
- Ben-Assuli, Ofir (2015). “Electronic health records, adoption, quality of care, legal and privacy issues and their implementation in emergency departments.” ENG. In: *Health policy (Amsterdam, Netherlands)* 119.3, pp. 287–297. ISSN: 1872-6054. DOI: 10.1016/j.healthpol.2014.11.014. URL: <http://www.ncbi.nlm.nih.gov/pubmed/25483873>.
- Boland, Michael V. et al. (2013). “Adoption of electronic health records and preparations for demonstrating meaningful use: an American Academy

- of Ophthalmology survey.” eng. In: *Ophthalmology* 120.8, pp. 1702–1710. ISSN: 1549-4713. DOI: 10.1016/j.ophtha.2013.04.029. URL: <http://www.ncbi.nlm.nih.gov/pubmed/23806425>.
- Bramble, James D. et al. (2010). “The relationship between physician practice characteristics and physician adoption of electronic health records.” eng. In: *Health care management review* 35.1, pp. 55–64. ISSN: 1550-5030. DOI: 10.1097/HMR.0b013e3181c3f9ad. URL: <http://www.ncbi.nlm.nih.gov/pubmed/20010013>.
- Buntin, Melinda Beeuwkes et al. (2011). “The benefits of health information technology: a review of the recent literature shows predominantly positive results.” eng. In: *Health affairs (Project Hope)* 30.3, pp. 464–471. ISSN: 1544-5208. DOI: 10.1377/hlthaff.2011.0178. URL: <http://www.ncbi.nlm.nih.gov/pubmed/21383365>.
- Burt, Catharine W. and Jane E. Sisk (2005). “Which physicians and practices are using electronic medical records?” eng. In: *Health affairs (Project Hope)* 24.5, pp. 1334–1343. ISSN: 0278-2715. DOI: 10.1377/hlthaff.24.5.1334. URL: <http://www.ncbi.nlm.nih.gov/pubmed/16162581>.
- Bush, George W. (2004). *Incentives for the Use of Health Information Technology and Establishing the Position of the National Health Information Technology Coordinator.*
- Chaudhry, Basit et al. (2006). “Systematic review: impact of health information technology on quality, efficiency, and costs of medical care.” eng. In: *Annals of internal medicine* 144.10, pp. 742–752. ISSN: 1539-3704. URL: <http://www.ncbi.nlm.nih.gov/pubmed/16702590>.
- Corrigan, JM et al. (2001). *Crossing the Quality Chasm: A New Health System for the 21st Century*. Washington, DC: National Academy Press.
- Decker, Sandra L., Eric W. Jamoom, and Jane E. Sisk (2012). “Physicians in nonprimary care and small practices and those age 55 and older lag in adopting electronic health record systems.” eng. In: *Health affairs (Project Hope)* 31.5, pp. 1108–1114. ISSN: 1544-5208. DOI: 10.1377/hlthaff.2011.1121. URL: <http://www.ncbi.nlm.nih.gov/pubmed/22535502>.
- Defining Key Health Information Technology Terms* (2008). Tech. rep. National Alliance for Health Information Technology.
- DesRoches, Catherine (2015). “Progress and challenges in electronic health record adoption: findings from a national survey of physicians.” eng. In: *Annals of internal medicine* 162.5, p. 396. ISSN: 1539-3704. DOI:

- 10.7326/L15-5060. URL: <http://www.ncbi.nlm.nih.gov/pubmed/25732297>.
- DesRoches, Catherine M. et al. (2008). "Electronic health records in ambulatory care—a national survey of physicians." eng. In: *The New England journal of medicine* 359.1, pp. 50–60. ISSN: 1533-4406. DOI: 10.1056/NEJMsa0802005. URL: <http://www.ncbi.nlm.nih.gov/pubmed/18565855>.
- DesRoches, Catherine M. et al. (2010). "Electronic health records' limited successes suggest more targeted uses." eng. In: *Health affairs (Project Hope)* 29.4, pp. 639–646. ISSN: 1544-5208. DOI: 10.1377/hlthaff.2009.1086. URL: <http://www.ncbi.nlm.nih.gov/pubmed/20368593>.
- DesRoches, Catherine M. et al. (2013). "Adoption of electronic health records grows rapidly, but fewer than half of US hospitals had at least a basic system in 2012." eng. In: *Health affairs (Project Hope)* 32.8, pp. 1478–1485. ISSN: 1544-5208. DOI: 10.1377/hlthaff.2013.0308. URL: <http://www.ncbi.nlm.nih.gov/pubmed/23840052>.
- Elnahal, Shereef M. et al. (2011). "Electronic health record functions differ between best and worst hospitals." eng. In: *The American journal of managed care* 17.4, e121–e147. ISSN: 1936-2692. URL: <http://www.ncbi.nlm.nih.gov/pubmed/21774097>.
- Ford, Eric W., Nir Menachemi, and M. Thad Phillips (2006). "Predicting the adoption of electronic health records by physicians: when will health care be paperless?" eng. In: *Journal of the American Medical Informatics Association : JAMIA* 13.1, pp. 106–112. ISSN: 1067-5027. DOI: 10.1197/jamia.M1913. URL: <http://www.ncbi.nlm.nih.gov/pubmed/16221936>.
- Ford, Eric W. et al. (2009). "Resistance is futile: but it is slowing the pace of EHR adoption nonetheless." eng. In: *Journal of the American Medical Informatics Association : JAMIA* 16.3, pp. 274–281. ISSN: 1067-5027. DOI: 10.1197/jamia.M3042. URL: <http://www.ncbi.nlm.nih.gov/pubmed/19261931>.
- Fredekind, Richard (2010). "Electronic health record conversion in a private dental school." eng. In: *The Journal of the American College of Dentists* 77.1, pp. 16–20. ISSN: 0002-7979. URL: <http://www.ncbi.nlm.nih.gov/pubmed/20481067>.
- Fricton, James et al. (2011). "The effect of electronic health records on the use of clinical care guidelines for patients with medically complex con-

- ditions.” eng. In: *Journal of the American Dental Association (1939)* 142.10, pp. 1133–1142. ISSN: 1943-4723. URL: <http://www.ncbi.nlm.nih.gov/pubmed/21965486>.
- Furukawa, Michael F. et al. (2014). “Despite substantial progress In EHR adoption, health information exchange and patient engagement remain low in office settings.” eng. In: *Health affairs (Project Hope)* 33.9, pp. 1672–1679. ISSN: 1544-5208. DOI: 10.1377/hlthaff.2014.0445. URL: <http://www.ncbi.nlm.nih.gov/pubmed/25104827>.
- Geisler, Benjamin P., Jeremiah D. Schuur, and Daniel J. Pallin (2010). “Estimates of electronic medical records in U.S. Emergency departments.” eng. In: *PloS one* 5.2, e9274. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0009274. URL: <http://www.ncbi.nlm.nih.gov/pubmed/20174660>.
- Ginn, Gregory O., Jay J. Shen, and Charles B. Moseley (2011). “Hospital financial position and the adoption of electronic health records.” eng. In: *Journal of healthcare management / American College of Healthcare Executives* 56.5, 337–50; discussion 351–2. ISSN: 1096-9012. URL: <http://www.ncbi.nlm.nih.gov/pubmed/21991681>.
- Goldzweig, Caroline Lubick et al. (2009). “Costs and benefits of health information technology: new trends from the literature.” eng. In: *Health affairs (Project Hope)* 28.2, w282–w293. ISSN: 1544-5208. DOI: 10.1377/hlthaff.28.2.w282. URL: <http://www.ncbi.nlm.nih.gov/pubmed/19174390>.
- Grinspan, Z. M. et al. (2013). “Physician specialty and variations in adoption of electronic health records.” eng. In: *Applied clinical informatics* 4.2, pp. 225–240. ISSN: 1869-0327. DOI: 10.4338/ACI-2013-02-RA-0015. URL: <http://www.ncbi.nlm.nih.gov/pubmed/23874360>.
- Hing, Esther and Catharine W. Burt (2009). “Are there patient disparities when electronic health records are adopted?” eng. In: *Journal of health care for the poor and underserved* 20.2, pp. 473–488. ISSN: 1049-2089. DOI: 10.1353/hpu.0.0143. URL: <http://www.ncbi.nlm.nih.gov/pubmed/19395843>.
- Hsiao C.-J., Burt C. W. Rechtsteniner E. Hing E. Woodwell D. & Sisk J. E. (2008). “Preliminary Estimates of Electronic Medical Record Use by Office-baed Physicians: United States, 2008.” In: *National Center for Health Statistics*.

- Hsiao, Chun-Ju and Esther Hing (2012). "Use and characteristics of electronic health record systems among office-based physician practices: United States, 2001-2012." eng. In: *NCHS data brief* 111, pp. 1–8. ISSN: 1941-4927. URL: <http://www.ncbi.nlm.nih.gov/pubmed/23384787>.
- Hsiao, Chun-Ju et al. (2011). "Electronic health record systems and intent to apply for meaningful use incentives among office-based physician practices: United States, 2001-2011." eng. In: *NCHS data brief* 79, pp. 1–8. ISSN: 1941-4927. URL: <http://www.ncbi.nlm.nih.gov/pubmed/22617322>.
- Imbens, Guido and Karthik Kalyanaraman (2011). "Optimal bandwidth choice for the regression discontinuity estimator". In: *The Review of Economic Studies*, rdr043.
- Jamoom, Eric and Esther Hing (2015). "Progress with electronic health record adoption among emergency and outpatient departments: United States, 2006-2011." eng. In: *NCHS data brief* 187, pp. 1–8. ISSN: 1941-4927. URL: <http://www.ncbi.nlm.nih.gov/pubmed/25714041>.
- Jha, Ashish K. et al. (2009a). "Electronic health records: use, barriers and satisfaction among physicians who care for black and Hispanic patients." eng. In: *Journal of evaluation in clinical practice* 15.1, pp. 158–163. ISSN: 1365-2753. DOI: 10.1111/j.1365-2753.2008.00975.x. URL: <http://www.ncbi.nlm.nih.gov/pubmed/18759752>.
- Jha, Ashish K. et al. (2009b). "Use of electronic health records in U.S. hospitals." eng. In: *The New England journal of medicine* 360.16, pp. 1628–1638. ISSN: 1533-4406. DOI: 10.1056/NEJMsa0900592. URL: <http://www.ncbi.nlm.nih.gov/pubmed/19321858>.
- Jha, Ashish K. et al. (2010). "A progress report on electronic health records in U.S. hospitals." eng. In: *Health affairs (Project Hope)* 29.10, pp. 1951–1957. ISSN: 1544-5208. DOI: 10.1377/hlthaff.2010.0502. URL: <http://www.ncbi.nlm.nih.gov/pubmed/20798168>.
- Jones, Spencer S. et al. (2010). "Electronic health record adoption and quality improvement in US hospitals." eng. In: *The American journal of managed care* 16.12 Suppl HIT, SP64–SP71. ISSN: 1936-2692. URL: <http://www.ncbi.nlm.nih.gov/pubmed/21314225>.
- King, Jennifer, Michael F. Furukawa, and Melinda B. Buntin (2013). "Geographic variation in ambulatory electronic health record adoption: implications for underserved communities." eng. In: *Health services research* 48.6 Pt 1, pp. 2037–2059. ISSN: 1475-6773. DOI: 10.1111/

- 1475-6773.12078. URL: <http://www.ncbi.nlm.nih.gov/pubmed/23800087>.
- Kohn, L.T., J. Corrigan, and M.S. Donaldson (2000). *To err is human: building a safer health system*. Washington: National Academy Press.
- Kokkonen, Erik W. J. et al. (2013). "Use of electronic medical records differs by specialty and office settings." eng. In: *Journal of the American Medical Informatics Association : JAMIA* 20.e1, e33-e38. ISSN: 1527-974X. DOI: 10.1136/amiajnl-2012-001609. URL: <http://www.ncbi.nlm.nih.gov/pubmed/23538721>.
- Lagrew, David C., Harris R. Stutman, and Laurie Sicaeros (2008). "Voluntary physician adoption of an inpatient electronic medical record by obstetrician-gynecologists." eng. In: *American journal of obstetrics and gynecology* 198.6, 690.e1-5; discussion 690.e5-6. ISSN: 1097-6868. DOI: 10.1016/j.ajog.2008.03.022. URL: <http://www.ncbi.nlm.nih.gov/pubmed/18538155>.
- Lee, Jinhyung, Yong-Fang Kuo, and James S. Goodwin (2013). "The effect of electronic medical record adoption on outcomes in US hospitals." eng. In: *BMC health services research* 13, p. 39. ISSN: 1472-6963. DOI: 10.1186/1472-6963-13-39. URL: <http://www.ncbi.nlm.nih.gov/pubmed/23375071>.
- Lehmann, Christoph U. et al. (2015). "Use of electronic health record systems by office-based pediatricians." eng. In: *Pediatrics* 135.1, e7-15. ISSN: 1098-4275. DOI: 10.1542/peds.2014-1115. URL: <http://www.ncbi.nlm.nih.gov/pubmed/25548325>.
- Li, Chenghui (2011). "Use of health information technology by office-based physicians: comparison of two contemporaneous public-use physician surveys." eng. In: *Perspectives in health information management / AHIMA, American Health Information Management Association* 8, 1f. ISSN: 1559-4122. URL: <http://www.ncbi.nlm.nih.gov/pubmed/22016672>.
- McCrary, Justin (2008). "Manipulation of the running variable in the regression discontinuity design: A density test". In: *Journal of Econometrics* 142.2, pp. 698-714.
- McCullough, Jeffrey S. et al. (2010). "The effect of health information technology on quality in U.S. hospitals." eng. In: *Health affairs (Project Hope)* 29.4, pp. 647-654. ISSN: 1544-5208. DOI: 10.1377/hlthaff.2010.0155. URL: <http://www.ncbi.nlm.nih.gov/pubmed/20368594>.

- Menachemi, Nir, Thomas L. Powers, and Robert G. Brooks (2011). "Physician and practice characteristics associated with longitudinal increases in electronic health records adoption." eng. In: *Journal of healthcare management / American College of Healthcare Executives* 56.3, 183–97; discussion 197–8. ISSN: 1096-9012. URL: <http://www.ncbi.nlm.nih.gov/pubmed/21714373>.
- Menachemi, Nir et al. (2007). "The influence of payer mix on electronic health record adoption by physicians." eng. In: *Health care management review* 32.2, pp. 111–118. ISSN: 0361-6274. DOI: 10.1097/01.HMR.0000267791.02062.3f. URL: <http://www.ncbi.nlm.nih.gov/pubmed/17438394>.
- Menachemi, Nir et al. (2009). "Use of health information technology by children's hospitals in the United States." eng. In: *Pediatrics* 123 Suppl 2, S80–S84. ISSN: 1098-4275. DOI: 10.1542/peds.2008-1755F. URL: <http://www.ncbi.nlm.nih.gov/pubmed/19088234>.
- Menachemi, Nir et al. (2012). "Market factors and electronic medical record adoption in medical practices." eng. In: *Health care management review* 37.1, pp. 14–22. ISSN: 1550-5030. DOI: 10.1097/HMR.0b013e3182352562. URL: <http://www.ncbi.nlm.nih.gov/pubmed/22016180>.
- Nakamura, Mari M. et al. (2010). "Electronic health record adoption by children's hospitals in the United States." eng. In: *Archives of pediatrics & adolescent medicine* 164.12, pp. 1145–1151. ISSN: 1538-3628. DOI: 10.1001/archpediatrics.2010.234. URL: <http://www.ncbi.nlm.nih.gov/pubmed/21135344>.
- Pallin, Daniel J. et al. (2010). "Health information technology in US emergency departments." eng. In: *International journal of emergency medicine* 3.3, pp. 181–185. ISSN: 1865-1380. DOI: 10.1007/s12245-010-0170-3. URL: <http://www.ncbi.nlm.nih.gov/pubmed/21031043>.
- Raglan, Greta B. et al. (2014). "Electronic Health Record Adoption among Obstetrician/Gynecologists in the United States: Physician Practices and Satisfaction." ENG. In: *Journal for healthcare quality : official publication of the National Association for Healthcare Quality*. ISSN: 1945-1474. DOI: 10.1111/jhq.12072. URL: <http://www.ncbi.nlm.nih.gov/pubmed/24779440>.
- Rao, Sowmya R. et al. (2011). "Electronic health records in small physician practices: availability, use, and perceived benefits." eng. In: *Journal of the American Medical Informatics Association : JAMIA* 18.3, pp. 271–

275. ISSN: 1527-974X. DOI: 10.1136/amiajn1-2010-000010. URL: <http://www.ncbi.nlm.nih.gov/pubmed/21486885>.
- Robert-Ribes, J (2005). "Predicting the Speed of Technology Introduction". In: *Austin Venture Capital Journal* 131, pp. 34–6.
- Rogers, Everett M. (1962). *Diffusion of innovations*. 4th. Free Press, New York, XVII, 519 s.
- (2003). *Diffusion of innovations*. New York: Free Press.
- Shields, Alexandra E. et al. (2007). "Adoption of health information technology in community health centers: results of a national survey." eng. In: *Health affairs (Project Hope)* 26.5, pp. 1373–1383. ISSN: 1544-5208. DOI: 10.1377/hlthaff.26.5.1373. URL: <http://www.ncbi.nlm.nih.gov/pubmed/17848448>.
- Shin, Dong Yeong et al. (2012). "Payer mix and EHR adoption in hospitals." eng. In: *Journal of healthcare management / American College of Healthcare Executives* 57.6, 435–48; discussion 449–50. ISSN: 1096-9012. URL: <http://www.ncbi.nlm.nih.gov/pubmed/23297609>.
- Simon, Steven R. et al. (2007). "Correlates of electronic health record adoption in office practices: a statewide survey." eng. In: *Journal of the American Medical Informatics Association : JAMIA* 14.1, pp. 110–117. ISSN: 1067-5027. DOI: 10.1197/jamia.M2187. URL: <http://www.ncbi.nlm.nih.gov/pubmed/17068351>.
- Simon, Steven R. et al. (2008). "Electronic health records: which practices have them, and how are clinicians using them?" eng. In: *Journal of evaluation in clinical practice* 14.1, pp. 43–47. ISSN: 1365-2753. DOI: 10.1111/j.1365-2753.2007.00787.x. URL: <http://www.ncbi.nlm.nih.gov/pubmed/18211642>.
- Singh, Ranjit et al. (2012). "The adoption and use of health information technology in rural areas: results of a national survey." eng. In: *The Journal of rural health : official journal of the American Rural Health Association and the National Rural Health Care Association* 28.1, pp. 16–27. ISSN: 1748-0361. DOI: 10.1111/j.1748-0361.2011.00370.x. URL: <http://www.ncbi.nlm.nih.gov/pubmed/22236311>.
- Spence, James M. et al. (2008). "Use of electronic forms within an electronic patient record to create an institutional clinical data repository." eng. In: *AMIA ... Annual Symposium proceedings / AMIA Symposium. AMIA Symposium*, p. 1146. ISSN: 1942-597X. URL: <http://www.ncbi.nlm.nih.gov/pubmed/18999032>.



- Stream, Glen R. (2009). "Trends in adoption of electronic health records by family physicians in Washington State." eng. In: *Informatics in primary care* 17.3, pp. 145–152. ISSN: 1476-0320. URL: <http://www.ncbi.nlm.nih.gov/pubmed/20074426>.
- Teufel, Ronald J. et al. (2013). "Electronic medical record adoption in hospitals that care for children." eng. In: *Academic pediatrics* 13.3, pp. 259–263. ISSN: 1876-2867. DOI: 10.1016/j.acap.2013.01.010. URL: <http://www.ncbi.nlm.nih.gov/pubmed/23680343>.
- The American Recovery and Reinvestment Act* (2009). The American Recovery and Reinvestment Act of 2009 (ARRA), Public Law No. 111-5, 123 Stat. 115 (2009 Feb 17).
- TMA Special Survey Electronic Medical Records Report* (2008). Tech. rep. Austin, TX: Texas Medical Association.
- TMA Special Survey Electronic Medical Records Report* (2009). Tech. rep. Austin, TX: Texas Medical Association.
- TMA Special Survey Electronic Medical Records Report* (2010). Tech. rep. Austin, TX: Texas Medical Association.
- TMA Special Survey Electronic Medical Records Report* (2012). Tech. rep. Austin, TX: Texas Medical Association.
- Whitacre, Brian E. (2015). "Rural EMR adoption rates overtake those in urban areas." ENG. In: *Journal of the American Medical Informatics Association : JAMIA*. ISSN: 1527-974X. DOI: 10.1093/jamia/ocu035. URL: <http://www.ncbi.nlm.nih.gov/pubmed/25665701>.
- Xierali, Imam M. et al. (2013). "The rise of electronic health record adoption among family physicians." eng. In: *Annals of family medicine* 11.1, pp. 14–19. ISSN: 1544-1717. DOI: 10.1370/afm.1461. URL: <http://www.ncbi.nlm.nih.gov/pubmed/23319501>.
- Yeager, Valerie A., Nir Menachemi, and Robert G. Brooks (2010). "EHR adoption among doctors who treat the elderly." eng. In: *Journal of evaluation in clinical practice* 16.6, pp. 1103–1107. ISSN: 1365-2753. DOI: 10.1111/j.1365-2753.2009.01277.x. URL: <http://www.ncbi.nlm.nih.gov/pubmed/21176000>.
- Zhang, Ning Jackie et al. (2013). "Health information technology adoption in U.S. acute care hospitals." eng. In: *Journal of medical systems* 37.2, p. 9907. ISSN: 0148-5598. DOI: 10.1007/s10916-012-9907-2. URL: <http://www.ncbi.nlm.nih.gov/pubmed/23340826>.

Zhivan, Natalia A. and Mark L. Diana (2012). "U.S. hospital efficiency and adoption of health information technology." eng. In: *Health care management science* 15.1, pp. 37–47. ISSN: 1386-9620. DOI: 10 . 1007 / s10729-011-9179-2. URL: <http://www.ncbi.nlm.nih.gov/pubmed/21922226>.