POTENTIAL COST BENEFIT FROM EARLY DETECTION OF CRITICAL CONGENITAL HEART DISEASE

By

JEREMY MICHAEL ARCHER

A THESIS PRESENTED TO THE GRADUATE SCHOOL OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE

UNIVERSITY OF FLORIDA

2013
To the children and families whose lives have been changed by congenital heart disease
ACKNOWLEDGMENTS

I would like to thank all those who made this work possible: Betsy Shenkman, for her encouragement and guidance throughout, providing access to a wonderful faculty and staff, and financial support; Barry Byrne for sparking my interest in newborn screening; Bruce Vogel for his patient teaching in health economics and econometric techniques; I-Chan Huang for his mentorship in patient-centered outcomes, Keith Muller for his assistance with the sample size calculation; Sarah Lynne-Landsman for reviewing an early draft of my proposal and suggesting methods to address selection bias; Melanie Sberna-Hinojosa and John Reiss for opening my eyes to the health disparities in my clinical field; Mildred Maldonado-Molina, Alex Wagenaar, and Kelli Komro for teaching fundamental research methodology; Jay Fricker and Arwa Saidi, for their support of my program of study; Katie Eddleton, for her assistance with preparing the Institutional Review Board protocol; Chris Shaffer, Chun He, Lei Zhang, Ashley Sanders, and Deepa Ranka for programming and data preparation; Jackie Hall, for geocoding; Doug Livingston, for an excellent discussion of multilevel modeling and intra-class correlation; Renata Shih, Mike Mattingly, and Srini Badugu for their support and encouragement; and the faculty and staff of both the Department of Health Outcome and Policy and the Division of Pediatric Cardiology for their teaching and other contributions. Finally, I would like to acknowledge and thank my wife, Jess Archer, for her support and patience during this work.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>4</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>8</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>9</td>
</tr>
<tr>
<td>LIST OF ABBREVIATIONS</td>
<td>10</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>11</td>
</tr>
<tr>
<td>CHAPTER</td>
<td></td>
</tr>
<tr>
<td>1 INTRODUCTION</td>
<td>13</td>
</tr>
<tr>
<td>Critical Congenital Heart Disease</td>
<td>14</td>
</tr>
<tr>
<td>Early and Late detection</td>
<td>14</td>
</tr>
<tr>
<td>Newborn Screening for CCHD</td>
<td>14</td>
</tr>
<tr>
<td>Racial/Ethnic Disparities in the Diagnosis of CHD</td>
<td>15</td>
</tr>
<tr>
<td>2 BACKGROUND</td>
<td>16</td>
</tr>
<tr>
<td>Scientific and Policy Context of Pulse Oximetry Screening for CCHD</td>
<td>16</td>
</tr>
<tr>
<td>Early Studies</td>
<td>16</td>
</tr>
<tr>
<td>Recent Studies</td>
<td>18</td>
</tr>
<tr>
<td>Technical Considerations</td>
<td>20</td>
</tr>
<tr>
<td>Saturation cutoff</td>
<td>20</td>
</tr>
<tr>
<td>Confirmation of abnormal saturation</td>
<td>21</td>
</tr>
<tr>
<td>Testing time</td>
<td>21</td>
</tr>
<tr>
<td>Pre- and post-ductal saturation</td>
<td>21</td>
</tr>
<tr>
<td>Oximetry equipment</td>
<td>22</td>
</tr>
<tr>
<td>Prenatal Diagnosis</td>
<td>22</td>
</tr>
<tr>
<td>Policy And Implementation</td>
<td>23</td>
</tr>
<tr>
<td>Previous Economic Analyses</td>
<td>23</td>
</tr>
<tr>
<td>Summary</td>
<td>24</td>
</tr>
<tr>
<td>Racial/Ethnic Disparities in CHD Care</td>
<td>25</td>
</tr>
<tr>
<td>Diagnosis</td>
<td>25</td>
</tr>
<tr>
<td>Overall Outcomes</td>
<td>26</td>
</tr>
<tr>
<td>Operative Timing and Quality</td>
<td>27</td>
</tr>
<tr>
<td>Neurodevelopmental Outcomes</td>
<td>29</td>
</tr>
<tr>
<td>Worldwide Access to Care</td>
<td>29</td>
</tr>
<tr>
<td>Summary</td>
<td>30</td>
</tr>
<tr>
<td>3 METHODS</td>
<td>32</td>
</tr>
</tbody>
</table>
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-1</td>
<td>Vulnerable populations in CHD</td>
<td>31</td>
</tr>
<tr>
<td>3-1</td>
<td>Outcome variables</td>
<td>47</td>
</tr>
<tr>
<td>3-2</td>
<td>Primary independent variable</td>
<td>47</td>
</tr>
<tr>
<td>3-3</td>
<td>Covariates</td>
<td>48</td>
</tr>
<tr>
<td>3-4</td>
<td>Internal variables</td>
<td>49</td>
</tr>
<tr>
<td>3-5</td>
<td>Algorithm for assigning early versus late detection</td>
<td>49</td>
</tr>
<tr>
<td>3-6</td>
<td>Diagnosis list A: congenital heart disease</td>
<td>50</td>
</tr>
<tr>
<td>3-7</td>
<td>Diagnosis list B: critical illness</td>
<td>50</td>
</tr>
<tr>
<td>3-8</td>
<td>Prematurity coding</td>
<td>52</td>
</tr>
<tr>
<td>3-9</td>
<td>Low birth weight coding</td>
<td>52</td>
</tr>
<tr>
<td>4-1</td>
<td>Missing data</td>
<td>63</td>
</tr>
<tr>
<td>4-2</td>
<td>Descriptive statistics for continuous data</td>
<td>63</td>
</tr>
<tr>
<td>4-3</td>
<td>Descriptive characteristics of categorical data</td>
<td>64</td>
</tr>
<tr>
<td>4-4</td>
<td>Final outcome model results</td>
<td>69</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>-----------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>3-1</td>
<td>Data workflow</td>
<td>47</td>
</tr>
<tr>
<td>3-2</td>
<td>Hospitals in Texas</td>
<td>51</td>
</tr>
<tr>
<td>4-1</td>
<td>Distribution of raw cost</td>
<td>65</td>
</tr>
<tr>
<td>4-2</td>
<td>Distribution of log-transformed cost</td>
<td>65</td>
</tr>
<tr>
<td>4-3</td>
<td>Comparison of log-transformed cost</td>
<td>66</td>
</tr>
<tr>
<td>4-4</td>
<td>Comparison of CRG</td>
<td>66</td>
</tr>
<tr>
<td>4-5</td>
<td>Comparison of interhospital distance</td>
<td>67</td>
</tr>
<tr>
<td>4-6</td>
<td>Park test for heteroskedasticity</td>
<td>68</td>
</tr>
<tr>
<td>4-7</td>
<td>Log-transformed cost by racial/ethnic category</td>
<td>70</td>
</tr>
<tr>
<td>4-8</td>
<td>Median income by racial/ethnic category</td>
<td>71</td>
</tr>
<tr>
<td>4-9</td>
<td>Percent of children in poverty in census tract by race/ethnicity</td>
<td>71</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td>CCHD</td>
<td>Critical congenital heart disease</td>
<td></td>
</tr>
<tr>
<td>CHD</td>
<td>Congenital heart disease</td>
<td></td>
</tr>
<tr>
<td>CHIP</td>
<td>Children’s Health Insurance Program</td>
<td></td>
</tr>
<tr>
<td>CRG</td>
<td>Clinical Risk Group</td>
<td></td>
</tr>
<tr>
<td>DHHS</td>
<td>Department of Health and Human Services</td>
<td></td>
</tr>
<tr>
<td>GLM</td>
<td>Generalized linear model</td>
<td></td>
</tr>
<tr>
<td>ICC</td>
<td>Intra-Class Correlation Coefficient</td>
<td></td>
</tr>
<tr>
<td>ICD-9</td>
<td>International Classification of Disease, 9th edition</td>
<td></td>
</tr>
<tr>
<td>ICHP</td>
<td>Institute for Child Health Policy</td>
<td></td>
</tr>
<tr>
<td>IRB</td>
<td>Institutional Review Board</td>
<td></td>
</tr>
<tr>
<td>IV</td>
<td>Instrumental variable</td>
<td></td>
</tr>
<tr>
<td>NBS</td>
<td>Newborn screening</td>
<td></td>
</tr>
<tr>
<td>OLS</td>
<td>Ordinary least squares</td>
<td></td>
</tr>
<tr>
<td>PHI</td>
<td>Protected health information</td>
<td></td>
</tr>
<tr>
<td>PI</td>
<td>Principal investigator</td>
<td></td>
</tr>
<tr>
<td>SES</td>
<td>Socioeconomic Status</td>
<td></td>
</tr>
<tr>
<td>UF</td>
<td>University of Florida</td>
<td></td>
</tr>
<tr>
<td>US</td>
<td>United States</td>
<td></td>
</tr>
</tbody>
</table>
POTENTIAL COST BENEFIT FROM EARLY DETECTION OF CRITICAL CONGENITAL HEART DISEASE

By

Jeremy Michael Archer

May 2013

Chair: Elizabeth Shenkman
Major: Medical Sciences – Health Outcomes and Policy

Introduction. CHD is the most common type of birth defect, and the largest cause of birth-defect related mortality. NBS using pulse oximetry is effective at detecting much CCHD, and detection of CCHD before initial hospital discharge has been associated with better short-term survival and other outcomes.

Methods. Using Medicaid and CHIP claims data from infants born in 2008-2009 in Texas, with follow-up data through 2011, we conducted a quasi-experimental study using a multilevel generalized linear modeling approach to test the relationship of timing of diagnosis with long-term cost of care, controlling for multiple explanatory variables. Both instrumental variable and treatment effects modeling were used to evaluate for group selection bias. We also evaluated the overall disease burden of both groups using the CRG score, and examined differences in race/ethnicity and SES between groups.

Results. Early detection of CCHD predicted a long-term health care cost reduction of 16%, or $18,000, when compared to late detection. Non-white race, extracardiac anomalies, aortic arch obstruction, and single ventricle physiology also predicted increased cost. More of those with late detection of CCHD were in the most
complex CRG category.

**Conclusion.** Early detection of CCHD predicts reduced health care cost, which could offset the increased cost associated with pulse-oximetry based NBS programs. Early detection may also be associated with increased overall disease burden and health care utilization. The association of non-white race with increased cost is unexplained.
CHAPTER 1
INTRODUCTION

CHD is the most common type of birth defect, affecting between 0.8 and 1% of the population, of which one third have "critical" CHD and require surgery or other intervention in the first month of life.\textsuperscript{1-4} CHD is also the single biggest cause of infant mortality attributable to birth defects.\textsuperscript{5-6} Moreover, neonates who are discharged into the community prior to diagnosis of their CCHD and then return to medical care have worse short-term outcomes, including preoperative condition and operative mortality.\textsuperscript{7-8} To investigate a potential means of avoiding delayed diagnosis, several large studies have shown pulse oximetry screening in the newborn nursery can diagnose most CCHD with reasonable accuracy.\textsuperscript{9-11} Indeed, universal screening programs have been implemented in several European countries and at least eight U.S. states, with pending legislation in 22 and pilot programs in seven others,\textsuperscript{12-13} and both the US DHHS and the American Academy of Pediatrics have recommended universal NBS with pulse oximetry.\textsuperscript{14} While there have been cost analyses of limited screening programs\textsuperscript{10-11, 15-18}, and cost-effectiveness models based upon anticipated program implementation\textsuperscript{19-21}, there is little data on the actual difference in health care cost from early diagnosis of CCHD, whether by a NBS program or through other means. We used the administrative claims databases of the Texas Medicaid and CHIP programs to compare paid amounts in the first four years of life between infants whose CCHD was detected before or during their birth hospitalization and those with delayed diagnosis of CCHD. Such a direct cost comparison has not been reported before. Although this study does not address a universal NBS program or pulse oximetry \textit{per se}, our hope is to provide insight into the direct effect of timely diagnosis on actual costs in a Medicaid/CHIP
population in a large and diverse state. We believe that this study will inform clinical and policy decisions as individual hospitals and states enact NBS programs for CCHD.

Secondary analyses included the overall disease burden as well as racial/ethnic and SES disparities in diagnosis and cost.

**Critical Congenital Heart Disease**

CCHD is defined as CHD requiring surgery or transcatheter intervention within the first 28 days of life, although some authors extend this window to one year and others use “serious,” “significant,” “major,” or “life-threatening” to represent the concept.\(^1\),\(^21\) The U.S. Department of Health and Human Services defines seven specific defects as CCHD: “hypoplastic left heart syndrome, pulmonary atresia (with intact septum), tetralogy of Fallot, total anomalous pulmonary venous return, transposition of the great arteries, tricuspid atresia, and truncus arteriosus,”\(^22\) focusing screening efforts on cyanotic disease\(^23\) despite clear evidence that left-sided obstructive disease, which is often non-cyanotic and therefore often missed on physical examination, may have worse outcomes if diagnosis is delayed.\(^8\),\(^24\)-\(^27\)

**Early and Late detection**

We defined early detection as diagnosis of CCHD prior to discharge from the birth hospitalization. This can occur antenatally or postnatally. We define late detection as diagnosis of CCHD after discharge from the birth hospitalization, or initial diagnosis if birth was not in a hospital setting.

**Newborn Screening for CCHD**

Those with late detection of CCHD typically have worse survival and physiologic outcomes than those with early detection, and this is particularly true in cases of left-sided obstructive disease, where the presentation following closure of the ductus
arteriosus can include acidosis, severe cardiogenic shock, and death.\textsuperscript{7-8, 24-27} Cyanosis is a common, although not universal, clinical feature of CCHD, and when mild may not be detectable on physical examination. Pulse oximetry offers a simple and relatively inexpensive adjunct to clinical examination as a potential screening tool for CCHD, and multiple studies have demonstrated that pulse oximetry combined with clinical examination has moderate sensitivity and high specificity with relatively few false-positives for detecting CCHD.\textsuperscript{9-12, 18, 28-38}

**Racial/Ethnic Disparities in the Diagnosis of CHD**

Important outcomes in CHD include diagnosis rates, survival, surgical timing and mortality, transplantation results, and neurocognitive development. While there has been significant investigation into the disparities found in adult cardiovascular care and outcomes, far less work has been done to investigate the determinants of and disparities in care and outcomes for children with CHD.\textsuperscript{39} Study of this topic is complicated by the fact that both overall and lesion-specific prevalence rates vary by both gender and race/ethnicity, so analyses must adjust for these poorly-understood but well-recognized differences.\textsuperscript{40-41} As a secondary aim of this paper, we will investigate the role of racial and ethnic disparities, in particular, in the timing of diagnosis and cost of care for children with CCHD.
CHAPTER 2
BACKGROUND

In this chapter, we provide a scientific and policy context for the move towards universal NBS for CCHD in the United States, as well as detailing some of its technical aspects. We then review the current status of health disparities research as it relates to the diagnosis and long-term outcomes of CHD.

Scientific and Policy Context of Pulse Oximetry Screening for CCHD

Early Studies

In the early 2000s, studies of pulse oximetry focused on feasibility and test performance characteristics. Hoke and colleagues performed pre- and post-ductal saturation measurements on 2876 infants, demonstrating a sensitivity of 85% for left-sided obstructive lesions and a sensitivity of 79% for other forms of disease, using a relatively low saturation cutoff of <92% and a wide pre-/post-ductal difference cutoff of ≥7%.\textsuperscript{28} They also showed that in left-sided obstructive disease, pre- and post ductal saturations differed while in cyanotic, CHD, the two measurements did not differ.

In Richmond and colleagues’ study of 6166 infants in a community hospital screening post-ductal saturation with a cutoff of ≤95% would have detected 30 of 40 (75%) cases of isolated congenital heart disease.\textsuperscript{29} Importantly, this study also demonstrated the utility of saturation screening in detecting non-cardiac disease as well. Kopell and colleagues used the same cutoff but screened at 24 hours of age or older, achieving a false positive rate of <0.001%.\textsuperscript{30} The low false positive rate was presumably because waiting until after 24 hours of life to screen allowed for a more complete transition to extra-uterine physiology, although it should be noted that their population sample also had a lower prevalence of CHD than has been published
elsewhere. Multiple other single-center studies in the U.S., Italy, Switzerland, Thailand, and Saudi Arabia confirmed the feasibility of implementing pulse oximetry as an adjunct to clinical examination, as well as its moderate to high sensitivity (67-100%), and high specificity (99-100%) when using post-ductal cutoffs of <95% or ≤95%.31-34

Meberg and colleagues reported the first multicenter study of pulse oximetry screening, in Norway, using a post-ductal saturation measurement on admission to the nursery with a cutoff of <95%.9 This study incorporated a repeat measurement to confirm abnormal results, demonstrating a sensitivity of 77.1%, a specificity of 94%, and a false-positive rate of 0.6%. As in prior studies,29-30 more than half of the abnormal screens detected non-cardiac disease, although this is likely reflected testing during the physiologic transition period from fetal to neonatal life. When this study was re-analyzed with non-screening regions of Norway as a control group, 12% of CCHD was missed in the screened group versus 23% in the control group; of the missed cases 82% were left-sided obstructive disease.36 There were no deaths from unrecognized CCHD, and cost was not studied.

A multicenter Swedish study combined pulse oximetry, with a post-ductal cutoff of ≤95% and a pre-/post-ductal difference cutoff of >3% and clinical examination to report a sensitivity of 82.8% and a specificity of 97.9% with a false positive rate of 2.1%.11 This study also compared results with a non-randomized control cohort, finding fewer deaths complications in the screened region, although the outcomes were too rare to draw conclusions. Unfortunately, both of these quasi-experimental approaches with non-randomized control groups failed to meaningfully address group selection bias.
in their design and analysis and so their inferences conclusions regarding outcome in the screened versus unscreened groups are relatively weak.

**Recent Studies**

At least one other multicenter study substantiated early results regarding the test performance characteristics of pulse oximetry,\textsuperscript{17, 42} although two studies of moderate size (15233 and 7672 patients) did show that the addition of routine pulse oximetry failed to yield any additional diagnoses of CCHD\textsuperscript{18, 43} Systematic reviews of early studies with pooled data analysis corroborated the moderate though variable sensitivity and high specificity found in early individual studies, and suggested further research\textsuperscript{44-45}, as did government-sponsored evidence reviews in the U.S. (Tennessee)\textsuperscript{46} and the U.K.\textsuperscript{19}

In the Middle Tennessee region, one such further study was performed after mandatory pulse oximetry screening was proposed in the state legislature.\textsuperscript{12, 46} Because screening was voluntary, seven of thirty hospitals, representing over half the births during the study period, did not participate. Of the 15,564 infants screened, 111 had an abnormal screen but only three were referred for further evaluation, a significant failure of the screening program. This pilot study utilized a cutoff of $<94\%$ with a single post-ductal measurement after 24 hours of age, with no repeat measurements, and points to the need for confirmatory repeat testing as well as a well-developed referral network. Tennessee has gone on to pass a law mandating universal pulse oximetry screening.\textsuperscript{13}

In a large multicenter study in Poland with 52,993 infants, Turska-Kmiec and colleagues demonstrated that a post-ductal measurement with a repeat confirmation of all positive results had good sensitivity and specificity, as well as nearly universal
acceptability to parents. Nearly 20% of CCHD was detected solely on the basis of the pulse oximetry screening alone in this study.

Bradshaw and colleagues demonstrated the feasibility of implementing a screening program with pre- and post-ductal measurement in a community hospital, reporting minimal additional cost and no additional staffing needs, a mean screening time of 3.5 minutes, and barriers to screening identified in only 2.4% of patients. The great majority of barriers reported related to screening equipment, staff workload, and crying or active newborns making the screening difficult.

In Great Britain, the landmark PulseOx study was a prospective evaluation of 20,055 newborns, in which all received both antenatal screening ultrasound and newborn pulse oximetry screening (pre- and post-ductal sites with ≤95% absolute and >2% difference cutoffs, with a repeat confirmation of abnormal screens). All infants were followed to 12 months of age to confirm diagnosis and outcome. Their findings included an overall sensitivity of 75% and specificity of 99%. This study was particularly valuable because it was done in the context of a rigorously measured antenatal detection rate of 50%, reflecting the added value of newborn pulse oximetry even with relatively high rates of prenatal diagnosis consistent with other areas. A related parent survey demonstrated that screening did not increase parental anxiety, although this effect was different among different ethnic groups.

Finally, Thangaratinam and colleagues performed a meta-analysis based upon all published studies including the PulseOx study, finding an overall sensitivity of 76.5% and specificity of 99.9%, with false positive rates of 0.05% when screening was conducted after 24 hours of life. They concluded based upon the evidence to date that
there is “compelling” evidence for routine pulse oximetry screening for CCHD. Based upon the extant evidence and an advisory committee recommendation,\(^\text{16}\) the US DHHS approved the implementation of pulse oximetry as part of the NBS panel routinely performed for all newborn infants.\(^\text{54}\)

**Technical Considerations**

As the studies of pulse oximetry have highlighted, there are at least five important technical considerations involved in testing: the saturation cutoff, confirmatory measurement for abnormal tests, testing time, the measurement of a pre- and post-ductal saturation difference, and the oximetry equipment itself.

**Saturation cutoff**

Early studies used a variety of saturation cutoffs, ranging from <92% to ≤95% in the single saturation or post-ductal measurement, and from >3% to >7% for the pre-/post-ductal difference. In 2005, de-Wahl Granelli and colleagues measured saturation at >12 hours of life in 66 newborns with CCHD and 200 normal newborns, to compare test performance characteristics using various cutoffs. They found that using a cutoff of ≤95% or a difference of >3% between pre- and post-ductal measurements using “new-generation” oximeters achieved the best overall accuracy, with sensitivity of 98.5% and specificity of 96.0%, an positive predictive value of 88.9%.\(^\text{55}\) Unfortunately this study did not report receiver-operator curves, limiting analysis to cutoffs previously published in the literature. Still, most recent studies have used a post-ductal cutoff of < or ≤95% and a pre-/post-ductal difference cutoff of >2-3%.\(^\text{9-11, 38, 47}\)

Testing at higher altitudes may require the development of different saturation cutoffs.\(^\text{15, 56}\)
Confirmation of abnormal saturation

Repeat measurements were introduced in the Swedish study\textsuperscript{9}, and have been incorporated into most large studies since\textsuperscript{10-11}. Confirmatory measurements are felt to increase test accuracy and reduce the potential for false positive results. In fact, the US DHHS advisory committee report recommends two separate confirmations of saturations between 90 and 95\%, stressing that in infants with saturation of <90\% evaluation should not be delayed for repeat measurement.\textsuperscript{16}

Testing time

Although one study found that at least six minutes of testing was necessary for maximum reliability\textsuperscript{18}, others have demonstrated the feasibility and low cost of testing lasting 1-5 minutes\textsuperscript{16}, and the average testing time in the PulseOx study was 3.5 minutes.\textsuperscript{10}

Pre- and post-ductal saturation

A difference between pre- and post-ductal saturation (typically measured on the right hand and right foot) may reflect differential perfusion due to left-sided obstructive disease, which may not necessarily be cyanotic.\textsuperscript{10-11, 28, 31-32} The effort to detect pre- and post-ductal differences and is clearly worthwhile, although focusing on relative pulse amplitude (reflecting peripheral perfusion) may be a better approach.\textsuperscript{55, 57} Considering that even the earliest studies in this area focused on the severe consequences and relatively high rate of non-detection of left-sided obstructive disease,\textsuperscript{8, 11, 24-28, 58} the DHHS decision to define CCHD purely on the basis of cyanotic disease is puzzling.\textsuperscript{22-23}
Ruegger and colleagues\textsuperscript{59} studied the utility of the left hand as a preductal site, and found it useful in normal infants, but there were no infants with CHD included. The ear and nose are also candidates for pre-ductal measurements.

**Oximetry equipment**

Pulse oximetry technology has evolved significantly over the last decade. New-generation oximeters are more accurate and have less measurement error than older-generation or conventional oximeters.\textsuperscript{11, 57} Studies of screening programs should ensure uniformity of oximetry equipment, and implementation of screening programs should, ideally, use the new-generation equipment.

**Prenatal Diagnosis**

A major factor influencing the accuracy, cost-effectiveness, and importance of NBS for CCHD is prenatal diagnosis, usually in the form of obstetric screening ultrasound with referral for fetal echocardiography when available.

In the United States, prenatal diagnosis rates for critical or significant CHD range from 31.7 to 50\%,\textsuperscript{49-51} although higher rates are possible.\textsuperscript{52} In the PulseOx study, in which every participant received a screening obstetric ultrasound leading to fetal echocardiography if warranted, 50\% (12 of 24) of cases of CCHD were prenatally diagnosed.\textsuperscript{10}

Peiris and colleagues found that both increasing socioeconomic status and private health insurance were associated with higher likelihood of having received prenatal diagnosis.\textsuperscript{51}

All of these results point to the fact that improved NBS still has a major role to play, given that fetal ultrasound is neither highly accurate nor universally available for the prenatal diagnosis of critical congenital heart disease.
Policy And Implementation

Both the American Heart Association and the American Academy of Pediatrics have endorsed the US DHHS recommended the addition of CCHD to the recommended uniform screening panel.\textsuperscript{14, 54, 60} Screening, even in the absence of a state mandate, is being performed in many hospitals; in Wisconsin, for example, 28% of hospitals representing 35% of births in 2010 were performing routine pulse oximetry screening. As of this writing, at least nine U.S. States have passed legislation to mandate statewide screening programs, although the Centers for Disease Control report that the majority of these are still in the planning stages and face significant financial obstacles.\textsuperscript{13, 61}

A survey in 2007 demonstrated that just over half of responding U.S. pediatric cardiologists supported mandated pulse oximetry screening, but a more recent survey in light of new evidence and recommendations has not been repeated.\textsuperscript{62}

Surveys have evaluated the degree of implementation of routine pulse oximetry screening in other countries as well. A Swiss survey conducted in 2008, 3 years after recommendation of universal screening in that country, demonstrated that 76% of maternity units representing 85% of newborns performed routine pulse oximetry screening.\textsuperscript{63} In the United Kingdom in 2010, 7% of neonatal units reported the practice and in 2012, 18% did.\textsuperscript{64-66}

Previous Economic Analyses

Typical cost estimates for pulse oximetry screening range from $0-$10 per test\textsuperscript{10-11, 15-17}. However, Reich and colleagues, in their study of non-tertiary care hospitals in Florida, found a per-test cost of $1,410-$2,128 per six minute screen by a Licensed Practicing Nurse, the minimum time and level of training associated with the best test
performance. Although their cost model may have been over-inclusive, their very different estimate points to the possibility that actual per-test costs may vary significantly.

Using a decision analytic model to predict cost-effectiveness in Great Britain, Knowles, Griebsch, and colleagues found an incremental cost effectiveness ratio for pulse oximetry screening program of $7,500 (in 2009 currency assuming an exchange rate of USD$0.65 per British pound) per additional timely diagnosis of CCHD. When Ewer and colleagues later adapted this decision analytic model using the test performance characteristics from the PulseOx study, this incremental cost-effectiveness ratio was between $10,000 and $42,300 per additional timely diagnosis, depending upon model assumptions and the inclusion of prenatal diagnosis. This model uses typical per-test cost figures, but could certainly overestimate cost-effectiveness if per-test cost is underestimated.

A rigorous economic analysis in the United States has not been reported, although this would certainly advisable as programs are being implemented on a state-by-state basis. Several authors point out that CCHD is more prevalent than many other diseases recommended for routine NBS in the U.S. The cost is substantially less than that of newborn hearing screening, the prototype for NBS using a non blood-spot methodology. Further cost analysis will be necessary to determine the difference in long-term or lifetime health-care cost of those with CCHD detected by screening versus those whose disease is not diagnosed in a timely fashion.

**Summary**

Multiple studies have demonstrated the effectiveness of routine pulse oximetry as an adjunct to clinical examination to screen for CCHD. Given the US DHHS
recommendation for implementation and a number of calls for universal screening, and the fact that at least eight states have already mandated it\textsuperscript{13,67-69}, pediatricians, cardiologists, and others need to focus careful attention on proper implementation. This includes using an appropriate saturation cutoff, incorporation of pre-ductal saturation measurement, adequate testing time, standardized and modern equipment, and the availability of both echocardiography and timely referral. Although the current recommendation focuses on cyanotic disease, we should not abandon the goal of developing more effective ways to detect left-sided obstructive disease, whether prenatally or by NBS.

**Racial/Ethnic Disparities in CHD Care**

Important domains in the care of patients with CHD include diagnosis, overall outcome, the timing and quality of operative intervention, and longer-term neurodevelopmental outcomes. All of these factors have the potential to affect both cost and quality of life in children with CHD, and all are vulnerable to health disparity. In this paper, we adopt the Institute of Medicine's definition of health disparity, "racial or ethnic differences in the quality of health care...not due to access-related factors...or appropriateness of intervention,"\textsuperscript{70} although we remain cognizant of broader definitions incorporating access to care.\textsuperscript{71}

**Diagnosis**

Diagnosis of CHD is a prerequisite to the surgical and medical therapies that have dramatically decreased mortality over the last several decades, and thus important to outcomes.\textsuperscript{72} Diagnosis can occur at many points of care: before birth (prenatal or fetal diagnosis), during the birth hospitalization, after an infant is discharged home but during childhood, during adulthood, and after death on autopsy. In the U.S., the rate of
screening in both the prenatal and newborn periods is increasing, but by no means universal. Additionally, timing of referral to the pediatric cardiologist or the pediatric cardiology center directly affects diagnosis.

Fetal or prenatal diagnosis is acknowledged to be widely disparate in terms of access, technique, and rate of diagnosis of CHD. Peiris and colleague studied the interaction of race, socioeconomic quartile, neighborhood poverty level, and prenatal diagnosis in a hospital with an extremely high prenatal diagnosis rate. They found that although the rate of prenatal diagnosis was correlated with all economic predictor variables, only private health insurance was a strong predictor (odds ratio 3.7) of prenatal diagnosis when all variables were modeled together.

Because universal NBS for CCHD is in the early stages of implementation, and not all CHD is critical, referral by pediatricians and family physicians for diagnosis by a pediatric cardiologist is the next important time point in diagnosis. In 1993, Fixler and colleagues reported population-based data from Dallas County, Texas, showing that race/ethnicity and socioeconomic status were unrelated to average age at postnatal diagnosis for nine of the most common lesions. Perlstein and colleagues, in 1997, reported that non-urban location was associated with later referral to a pediatric cardiologist in the neonatal period, while after the neonatal period non-urban location, and managed care insurance were associated with later referral. This study did not include race/ethnicity in its analysis. So far, the evidence points to a rural/urban disparity, but not necessarily one related to race or socioeconomic status.

**Overall Outcomes**

A large review of U.S. death certificates from 1979-1997 reveals that although population-based mortality rates from CHD have declined from 2.5 to 1.5 per 100,000, a
39% decrease, the approximately 20% mortality gap between blacks and whites has not changed to any significant degree.\textsuperscript{72}

Nembhard and colleagues, in two separate large studies with population-based denominators, demonstrated higher rates of childhood and early childhood mortality for nonwhites relative to whites from CHD across several lesion types\textsuperscript{75-76} In some subgroups and in the overall analysis of the 2008 paper, Hispanic males and females had lower mortality than whites.\textsuperscript{75}

However, the picture of racial disparity in CHD is not all straightforward. Recently, the Centers for Disease Control reported on 2,256 neonatal deaths attributable to CHD, out of 11 million live births from 2003-2006.\textsuperscript{5} A higher proportion of neonatal deaths in children of white mothers (5.4%) were due to CHD versus those in children of black mothers (2.3%) In preterm infants, the neonatal death rate due to CHD was lower for children of black versus white mothers (4.5 versus 6.8 per 10,000), while in term births, neonatal mortality rates were higher for children of black versus white mothers (1.5 versus 1.0 per 10,000). These data illustrate the complexity of analyzing just three factors: race, prematurity, and neonatal death.

**Operative Timing and Quality**

Age at operation has been used as an indicator of both quality and access to care in several studies of health disparities in CHD.\textsuperscript{77-79} Additionally, hospital-based surgical mortality is an important quality indicator for pediatric cardiovascular surgery centers.

Erikson and colleagues analyzed California discharge data in 1992-1994 for 5071 patients undergoing congenital heart surgery, finding that those with managed care insurance had a lower chance of having an operation at a hospital with lower
Interestingly, after stratification by race/ethnicity, there was no difference in access to low-mortality hospitals between Medicaid and traditional private insurance, although the difference between managed care traditional private insurance persisted.

Chang and colleagues also examined statewide data in California in 1995-1996, finding that commercial (private) versus public insurance was associated with earlier repair of four specific defects, while urban location was associated with later repair.

Race had little to no impact on age of repair, except that Asian children underwent later repair.

These results again illustrate the complexity of untangling the factors influencing diagnosis and referral. One could theorize that urban children tend to have predominantly public insurance or some other latent factor related to less access or lower socioeconomic status, and are thus repaired later; one could also theorize that cardiologists simply referred these children for surgery later because they were close by and easier to watch medically. Similarly, there may be cultural or economic factors common to the “Asian” children (itself a heterogeneous group) in this study leading to later referral or later family consent to surgery.

In 2002, Milazzo and colleagues reported a relatively small, single-center study of timing of second and third-stage palliative surgeries for single ventricle physiology performed from 1997-2000. They found that African-American children underwent both surgeries significantly later than their white counterparts, by a difference of six months for the cavopulmonary anastomosis (Glenn) and two years for the Fontan completion surgery. However, when a similar study was repeat using data from 29
hospitals and a much larger sample size, no difference in median age was found, although Hispanic patients did have a longer length of ICU stay after one of the procedures.  

**Neurodevelopmental Outcomes**

The study of neurodevelopmental outcomes is at the forefront of pediatric cardiology, especially in the current era, in which most patients with CHD are expected to survive. Unfortunately, many of the landmark trials in this area have focused almost exclusively on physiologic or surgical factors, with less attention to health disparities or the influence of socioeconomic or demographic factors. This has been the case even when the focus is on social determinants of behavioral issues.

Atallah and colleagues did address socioeconomic status in a landmark study examining outcomes of two palliative surgery techniques for hypoplastic left heart syndrome, used in two historically different time periods. They found that socioeconomic status was a predictor of lower mental development scores with one technique, used in the earlier time period, although not with the other, used in the later time period. Male gender, on the other hand, was associated with lower psychomotor development scores in the more recent surgical time period.

**Worldwide Access to Care**

Finally, although this portion of the current study focuses primarily on the disparity landscape in the United States, worldwide disparities in access to pediatric cardiovascular care are significant. There is a gross mismatch in pediatric cardiac surgery centers relative to the birth rate of infants with CHD, and a call has been made for worldwide access to improve such access in conjunction with the United Nations 2000 Millennium Development Goals.
Summary

Current evidence reveals multiple disparities in the care and outcomes of patients with CHD, disparities that cut across race, gender, socioeconomic status, and geographic location. Groups identified as vulnerable by the current evidence are summarized in Table 2-1.

While several disparities have been shown in large studies and replicated multiple times, particularly the racial disparities in overall and postoperative survival, others require further and focused study. To this end, we include a descriptive analysis of racial/ethnic disparity in our study.
<table>
<thead>
<tr>
<th>Process or Outcome</th>
<th>Vulnerable Group(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prenatal Diagnosis</td>
<td>Publicly Insured</td>
</tr>
<tr>
<td>Postnatal Diagnosis</td>
<td>Rural</td>
</tr>
<tr>
<td>Access to Care</td>
<td>Developing World</td>
</tr>
<tr>
<td>Overall Mortality</td>
<td>Non-white</td>
</tr>
<tr>
<td>Age at Operation</td>
<td>Publicly Insured</td>
</tr>
<tr>
<td></td>
<td>Urban</td>
</tr>
<tr>
<td></td>
<td>Asian and African-American</td>
</tr>
<tr>
<td>Operation at Low-Mortality Center</td>
<td>Managed Care Insured</td>
</tr>
<tr>
<td>Neurocognitive Outcomes</td>
<td>Low SES</td>
</tr>
</tbody>
</table>
CHAPTER 3
METHODS

The study protocol was approved by the UF Health Science Center IRB (protocol IRB201200064) on November 30, 2012. A waiver of informed consent was granted. Neither the author nor any study participant declared any potential conflict of interest related to the study. The SAS System version 9.3 was used for all statistical analysis except where noted otherwise.

Theoretical Framework

The theoretical framework for our study was based upon the current literature showing that late detection of CCHD can lead to increased physiologic derangement of multiple organ systems, which could lead to differential outcomes. We included additional explanatory variables addressing race/ethnicity, SES, specific lesion physiology, additional birth defects, and birth-related factors, each of which could potentially influence group selection, outcomes, or both.

Specific Aims

Based upon the review of existing literature, we chose to focus on the following questions for this study:

1. Do children with early detection of CCHD have a different health care cost in the first four years of life when compared with those with late detection of CCHD?

2. What other measurable variables affect health care cost in these groups?

3. How do differences in race/ethnicity and SES affect the timing of detection and the health care cost in the first four years of life?

4. Compared with infants with late detection of CCHD, do infants with early detection have a different overall disease burden affecting health care utilization?
5. Compared with infants with late detection of CCHD, do infants with early detection have increased survival in the first five years of life?

Our goal was to address gaps in the literature using a methodology that rigorously evaluated and controlled for selection bias, heteroskedasticity, and clustering effects.

**Primary Hypothesis**

In children with CCHD enrolled in Medicaid and CHIP and born in Texas in 2008 and 2009, we hypothesized that those whose CCHD was detected prior to initial hospital discharge will have a lower health care cost in the first four years of life when compared with those whose CCHD was detected after initial discharge. (Hypothesis 1)

We did not distinguish between costs due directly to CCHD and costs due to other causes, as our theoretical framework describes the impact of CCHD on patients’ overall medical condition and cost.

As a complementary analysis, we hypothesized that there would be relationship between cost and multiple variables known to affect outcome in CHD or neonatal survival: gender, multiple gestation, delivery via Cesarean section, prematurity, low birth weight, the presence of an extracardiac birth defect, race/ethnicity, and SES as reflected in the median annual income and percent of children living below poverty by census tract or zip code midpoint. (Hypothesis 2)

**Secondary Aims**

**Racial/ethnic and SES disparity**

We hypothesized that there would be a significant relationship between race/ethnicity and cost, possibly via a role in group selection. (Hypothesis 3A)

Furthermore, we hypothesized that there would be a significant relationship between SES and cost of care, possibly via a role in group selection. (Hypothesis 3B)
Overall disease burden

We hypothesized that the overall disease burden affecting health care utilization, as measured in the pediatric CRG score, would be lower in those with early detection of CCHD versus those with late detection. (Hypothesis 4)

Survival

We hypothesized that the overall survival would be higher in those with early detection of CCHD versus those with late detection. (Hypothesis 5). This expected result would be consistent with prior published quasi-experimental work.\textsuperscript{11, 36}

Research Design

This is a quasi-experimental, retrospective design:

\begin{tabular}{c c c}
NR & X & O \\
\hline
NR & O \\
\end{tabular}

“NR” indicates non-randomized group selection, in this case early versus late diagnosis of CCHD. “X” indicates early diagnosis. “O” indicates the primary outcome measure, paid claims from birth through December 31, 2011.

To perform the analysis, we used a multilevel GLM, preceded by prospective determination of sample size, testing for heteroskedasticity, and evaluation for selection bias.

To perform secondary analyses, we adapted the same general design as well as using descriptive and simple comparative statistics for alternative outcomes.

Data and Variable Specifications

Data Sources

The state of Texas provides health claims, encounter, and enrollment data at the person level for its Medicaid and CHIP programs on a quarterly basis. This includes
Fee For Service, and Primary Care Case Management, and Health Maintenance Organization (STAR) structures. The Medicaid and CHIP enrollment and claims data are housed by UF for quality assurance activities carried out by the ICHP pursuant to its quality assurance contract with Texas (Evaluating Healthcare Quality in Texas Medicaid and CHIP, UF Project Number 0068240, Elizabeth Shenkman, PI). Elizabeth Shenkman, Ph.D., is the Professor and Chair of the Department of Health Outcomes and Policy and the Director for the ICHP.

**Research Subjects**

**Birth cohort**

The research subjects consisted of all babies born in 2008-2009 in TX who met our diagnostic criteria for CCHD, and were enrolled in Medicaid, of any type or CHIP at the time of birth.

**Follow-up data**

The data analyzed consisted of all Medicaid and CHIP enrollment and claims data for all children in the birth cohort, from birth through December 31, 2011.

**Data Workflow and Security**

The original databases, described above, reside on a secure server in the ICHP, under contracts approved by the state of Texas and approved by the UF IRB. These databases are maintained under strict security and data is not permitted to leave the server. For our study, the only study team members with access to the original data were the programming and geocoding team members. The programmers extracted all operational variables to an operational dataset at the individual study participant level. The geocoder used address and zip code information to extract income and poverty variables from American Community Survey data, and hospital distance variables from
the American Hospital Association data, and these were re-merged with the operational dataset. At this point, all PHI was removed from the dataset to form the de-identified analysis dataset. The de-identified analysis dataset was accessible only to the primary author (JA) who performed all data analysis and did not have access to the original databases or any PHI-containing data for this study. The operational workflow is summarized in Figure 3-1.

**Variable Definitions**

Seventy-one variables were operationalized into outcome variables, the primary independent variable, covariates, and internal variables as described in Tables 3-1, 3-2, 3-3, and 3-4, respectively

**Determining early versus late detection**

The algorithm used to determine early or late detection is given in Tables 3-5, 3-6, and 3-7. Infants born at home or out of the hospital setting and diagnosed with CCHD were automatically considered to have late detection.

**Geocoding**

Three variables were obtained using geocoding with ArcMap 10.1 using premium street data. Addresses were mapped to census tract, and the median income and percent of children living in poverty in a patient’s census tract were determined using the American Community Survey dataset from 2010. Because calculating income and poverty variables required census tract level information, subjects with no address, an address outside of Texas, a Post Office Box, or a zip code only were not assigned values for these variables.

The interhospital distance was calculated by subtracting the driving time in minutes of the patient’s address to the nearest hospital, regardless of the actual birth
hospital, from the driving time in minutes from the patient’s address to the nearest hospital that reported providing pediatric cardiac surgery services in the American Hospital Association survey for fiscal year 2007\(^7\) (see Figure 3-2). For interhospital distance, subjects with no address or an address outside of Texas were not assigned a value but subjects with a zip code only or a Post Office Box were assigned a value calculated using the midpoint of their zip code instead of a street address.

**Covariates**

Covariates were chosen based upon established risk models for neonatal survival and disease severity, \(^91-93\) as well as upon studies of factors influencing CCHD outcomes\(^8, 88-90\). Aortic arch obstruction, and in particular isolated coarctation, are sometimes difficult to diagnose after birth and are more common in those with delayed diagnosis. Single ventricle physiology is typically associated with a minimum of two to three palliative surgeries and so impacts both survival and cost. The birth hospital was used not strictly as a covariate, but as a clustering variable, as discussed below. The interhospital distance, similarly, was used as an instrumental variable to test and control for selection bias due to unmeasured variables. Finally, the total number of months enrolled during the study period was used as a covariate after log-transformation. The operational specifications of all covariates are given in Table 3-3, with diagnosis lists for prematurity and low birth weight given in Tables 3-8 and 3-9.

**Sample Size Determination**

A preliminary dataset consisting of infants born and enrolled in Texas Medicaid (all types) in 2011, was prepared for to estimate the sample size needed for the full analysis. The online program GLIMMPSE\(^99\) based in part upon work by Creidler, Glueck, and Muller,\(^100\) can calculate power and sample size for multilevel data in a
generalized linear model. We calculated the descriptive statistics, ICC, and the harmonic mean of the cluster size of the preliminary data, and used the GLIMMPSE program with the default Hotelling-Lawley Trace method to prospectively calculate an initial sample size. We then repeated the calculation assuming half of the difference in outcome variable and twice the variance as seen in the preliminary data to compute our final desired sample size.

After the final dataset was prepared, we retrospectively calculated another sample size based upon the actual characteristics of the data, to detect a difference in outcome half as large as was actually observed. This was used to confirm that our actual sample size was adequate.

**Preliminary Analysis**

**Missingness**

The percentage of all variables missing was determined. For variables with more than 5% missingness, a chi-squared test was used to determine if the degree of missingness was significantly different between the early and late selection groups. Missing values were not filled in or imputed, and observations with missing values were dropped from the final analysis.

**Normality Testing**

To determine if cost would satisfy the assumption of normality, it was analyzed using the SAS univariate procedure to determine if the variable fit a standard normal distribution using visual analysis of a histogram as well as the kurtosis and skewness. This procedure was repeated with the natural logarithm of cost (“log-transformed”).
**Descriptive Statistics**

For continuous variables, the SAS `ttest` procedure was used to compare means between the early and late detection groups. Similarly, the SAS `freq` procedure was used to determine the overall proportion of subjects with each value of categorical variables, and then to compare proportions between detection groups using the chi-square test.

**Intra-class Correlation**

Intra-class Correlation ICC for clustered data is given by

\[
ICC = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_w^2} \tag{3-1}
\]

where \(\sigma_b^2\) is the between-group variance and \(\sigma_w^2\) is the within-group variance. For this study, the SAS `mixed` procedure was used to determine the within- and between-group variances and to calculate the ICC. This was performed once while ignoring subjects with a missing birth hospital, and repeated with all subjects having a missing birth hospital treated as if they were in a single cluster.

**Park Test for Heteroskedasticity**

Manning and Mullahy\(^{101}\) describe the Park test for heteroskedasticity in log-transformed models of health care expenditures. This test regresses log-transformed predicted value against a log-transformed residual uses the estimated coefficient to choose the distribution function used in the final model. The simplified form of the procedure described by Manning and Mullahy\(^{101}\) is given by

\[
y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \tag{3-2}
\]

\[
\ln(y_i - \hat{y}_i)^2 = \lambda_0 + \lambda_1 \ln(\hat{y}_i) + v_i \tag{3-3}
\]
where (3-2) is either a GLM of a raw cost with a log link or an OLS model of log-transformed cost, and (3-3) is a linear regression, with the predicted value \( \hat{y}_i = \exp(x_i\hat{\beta}) \) if a GLM was used initially, or \( \hat{y}_i = \exp(x_i\hat{\beta} + 0.5\hat{\sigma}^2(x)) \) to correct for retransformation bias if an OLS model was used initially. The value of \( \lambda_1 \) determines the probability distribution for the final GLM: if \( \lambda_1 = 0 \) there is no appreciable heteroskedasticity and a Gaussian distribution is appropriate; if \( \lambda_1 = 1 \), Poisson is used; if \( \lambda_1 = 2 \), a gamma distribution is used; if \( \lambda_1 = 3 \), an inverse Gaussian distribution is used.

In this study, the Park test was performed three times. The first was with an OLS regression using the SAS `glm` procedure with a the appropriate correction for retransformation bias; the second was with a GLM using the SAS `genmod` procedure and a Poisson distribution, and the third was with a GLM again using the SAS `genmod` procedure and a gamma distribution. In all three cases the SAS `reg` procedure was used to model equation (3-3) and determine \( \lambda_1 \). For the Park test, multilevel modeling was not used.

**Selection Bias and Endogeneity**

Selection bias refers generally to the nonrandom sampling of the population. One specific case of this is group selection bias, in which nonrandom samples of the population are selected into groups for analysis. Although these are often referred to as treatment groups, we will use the term “selection group” in this study as we are not investigating a treatment. When we refer to selection bias, we mean group selection bias, although we do address the general issue of nonrandom sampling in Chapter 5. Because group selection was neither randomized nor under experimental control, there
was felt to be a high potential for group selection bias in our study, due to both measured and unmeasured variables.

Endogeneity refers to the correlation between the selection variable and the error term, which can incorporate unmeasured variables, measurement error, or dynamic features of the equation system. Selection bias and endogeneity are related, but not identical concepts, and endogeneity of the selection variable can be one reason for group selection bias.

In the absence of randomization, multiple approaches exist to detecting and controlling for selection bias and endogeneity. Many health outcomes researchers have used propensity scoring,\textsuperscript{102-105} in which the probability of selection into one group is determined based upon measured variables. The propensity score is then used to match the groups, as a control variable in the final model, or both. The disadvantage of this approach is that it fails to account for unmeasured variables, and may not control for endogeneity due to unmeasured variables.

**Instrumental Variable Analysis**

An alternative approach to endogeneity and selection bias, which accounts for both measured and unmeasured variables, is the use of instrumental variables. The instrumental variable, or instrument, should be strongly related to the selection variable and account for unmeasured variables in that relationship, and it should be otherwise independent of the outcome variable.\textsuperscript{106} Using the differential distance from the nearest hospital to a hospital offering specialty care as an instrument has been successful in previous neonatal\textsuperscript{106-107} and in the cardiovascular\textsuperscript{108-109} studies. For that reason, we chose differential distance to a pediatric cardiac surgery center as an instrument.
A standard instrumental variable approach involves predicting the probability of selection into a reference group based on the instrument and other variables. The predicted probability is typically used in place of the actual group selection variable in the final model, although it has been used for group matching as well.\textsuperscript{106} A standard instrumental variable model is represented by
\begin{align*}
w_i &= y_0 + y_1 z_i + y_2 x_i + \varepsilon_i \\
y_i &= \beta_0 + \beta_1 \hat{w}_i + \beta_2 x_i + u_i
\end{align*}
where $y$ is the outcome variable, $w$ is the group selection variable, $\hat{w}$ represents the predicted probability of being selected into a reference group, $z$ is the instrumental variable, and $x$ represents all additional exogenous covariates, some of which may influence group selection. In (3-5), note that the outcome variable is calculated based on the predicted probability, not the actual group selection variable.

Wooldridge\textsuperscript{110} describes a three stage instrumental variable approach given by
\begin{align*}
\phi_i &= \pi_0 + \pi_1 z_i + \pi_2 x_i + \nu_i \\
w_i &= y_0 + y_1 \hat{\phi}_i + y_2 x_i + \varepsilon_i \\
y_i &= \beta_0 + \beta_1 \hat{w}_i + \beta_2 x_i + u_i
\end{align*}
in which the predicted probability of selection into a reference group, $\hat{\phi}$, calculated in (3-6) is used as an instrument with the same covariates as in (3-7), producing a second predicted probability of selection into a reference group, $\hat{w}$, that is then used in the final outcome model (3-8). In both (3-6) and (3-7), the selection variable is the left hand side of the equation; in (3-8) it is the outcome variable. The primary advantages of this three-stage model are increased accuracy and efficiency if the model for $\hat{w}_i$ is misspecified.
Our instrumental variable approach involved computing predictor variables using both the two and the three stage methodology. The SAS logistic procedure was used to model the predictor variables. The strength and significance of the parameter estimate for the instrument in predicting the selection variable was examined to assess the strength of the instrument itself.

**Heckman Selection Model**

In addition to evaluating the suitability of the chosen instrument during the modeling process, we used the Heckman selection model\textsuperscript{111} as a treatment effects model to test the endogeneity of the selection variable and the potential for its influence on group selection. This involves treating both the selection an outcome variable as endogenous variables, simultaneously modeling a probit regression of the selection variable and a linear regression of the outcome variable\textsuperscript{112}, as given by

\[ \Pr(w_i = 1) = \gamma_0 + \gamma_1 z_i + \varepsilon_i \]  
\[ y_i = \beta_0 + \beta_1 w_i + \beta_2 x_i + u_i \]

Note that \( z \) represents a set of exogenous variables modeled with the group selection variable, while \( x \) represents a separate set of exogenous variables modeled with the outcome variable. A nonzero relationship between the disturbance terms \( \varepsilon_i \) and \( u_i \) is evidence of group selection bias. Adding exogenous variables (regressors) in the outcome equation that are not present in the prediction equation can make this approach more robust\textsuperscript{112}. The Heckman selection model simultaneously tests and controls for selection bias.
In our study we used the STATA\textsuperscript{113} \texttt{treatreg} command with a maximum likelihood model to perform a Heckman selection model. In this model, the $\rho$ statistic represents the correlation of disturbance terms.

**Modeling the Primary Hypothesis**

After assessing missingness, data normality, heteroskedasticity, and endogeneity and selection bias, we modeled our data using a multilevel generalized linear model with a log link. We ran multiple candidate models using all covariates and narrowed this down to a parsimonious model. Based upon the results of the preliminary analyses (presented in Chapter 4), we chose a gamma distribution function and did not include the instrumental variable in the final model. Because preliminary modeling did not find a significant relationship between months enrolled and cost, and because we did not use a Poisson distribution function in the final model, we included the log-transformed version of total months enrolled as a covariate rather than an offset. Detailed descriptions of candidate models, the final model, and alternative models are presented in Chapter 4 and Appendix A.

Although others have used mixed models to account for random effects,\textsuperscript{114} we were unable to achieve convergence with a generalized linear mixed model using the SAS \texttt{glimmix} procedure, presumably due to the complexity of the nonuniform clustering. A GLM or generalized estimating equation without random effects terms, using the SAS \texttt{genmod} procedure with an exchangeable covariance matrix, was found to be more robust and was used for all final data modeling.

The final model allowed us to test Hypothesis 1, while the progression from the maximal to parsimonious model allowed us to address hypothesis 2.
Estimates of change in cost by individual independent variables can be calculated using the parameter estimate for variable and the overall sample mean cost:

\[(e^{\beta_1} - 1) \cdot \bar{\text{cost}}\]  

(3-11)

Because a GLM is used there is no need to adjust for retransformation bias when re-exponentiating the parameter estimate \(\beta_1\).

Testing Secondary Aims

Racial/ethnic and SES Disparity

The continuous variables including cost, interhospital distance, income, poverty, and enrolled month were compared across race/ethnicity groups using analysis of variance with the SAS \texttt{anova} procedure. The categorical variables including gender, prematurity, low birth weight, multiple gestation, Cesarean section, aortic arch obstruction, aortic coarctation, single ventricle physiology, and clinical resource group were compared across race/ethnicity groups using chi-squared with the SAS \texttt{freq} procedure. To further assess for hidden or subtle disparities, these analyses were repeated comparing only two categories: white and non-white race/ethnicity. This was accomplished using the SAS \texttt{ttest} procedure for continuous variables and the SAS \texttt{freq} procedure for categorical variables. These analyses addressed Hypothesis 3A.

Finally, to assess the relationship of income and poverty to the outcome and the selection variables, both of these variables were included in several candidate models of the outcome variable. Additionally, their means were compared in the early versus late detection group using the SAS \texttt{ttest} procedure. Because of missingness, however, they were not included in the final model. These analyses addressed hypothesis 3B.
Overall Disease Burden

The CRG scoring system was developed to predict health care resource utilization\textsuperscript{94} and then adapted for pediatric use.\textsuperscript{95} It is readily calculable using claims data and may broadly reflect overall disease burden as it relates to healthcare utilization. We must be clear that it has not, to our knowledge, been used in this fashion before, and this is a preliminary investigation.

Our original intent had been to use maximum likelihood estimation will be used to compare the CRG score between detection groups, with a similar group of covariates as used in the final primary outcomes model. However, because of a high degree of missingness and concerns over the standardization of timing of the CRG given the different ages of the patients at the conclusion of the study, we opted instead to describe the differences in CRG between the early and late detection groups, analyzing it both as a continuous variable using the SAS \texttt{ttest} procedure and as a polychotomous, ordinal variable using chi-square with the SAS \texttt{freq} procedure. This addressed Hypothesis 4 but did not test it rigorously.

Survival

We originally planned to conduct a proportional hazards regression using the SAS \texttt{phreg} procedure. However, this analysis not performed due to incomplete and presumably inaccurate survival data. Hypothesis 5 was therefore not tested.
Table 3-1. Outcome variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Type</th>
<th>Operational Specification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost</td>
<td>Total paid amount in the first five years of life.</td>
<td>Continuous</td>
<td>Exp_Med: total Medicaid expenditures from birth through end of CY2011 ($0.00)</td>
</tr>
<tr>
<td>Comorbidity Index</td>
<td>Clinical Risk Group(^{95})</td>
<td>Ordinal</td>
<td>CRGSUM (code as integer): from CY2011 or most recent available year</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1: Healthy</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2: Significant Acute</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3: SHCN – Minor</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4: SHCN – Moderate</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>5: SHCN – Major</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Missing: Unassigned</td>
</tr>
<tr>
<td>Survival</td>
<td>Age at death (789.x), or survival to age 5</td>
<td>Continuous</td>
<td>Survival: survival days at death (the death date is as the first date with 789.x), or survival to December 31 2011 (integer)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>F_Death: flag of death (integer)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0=N</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1=Y</td>
</tr>
</tbody>
</table>

Table 3-2. Primary independent variable

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Type</th>
<th>Operational Specification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Detection Group</td>
<td>Early or late detection of CCHD.</td>
<td>Dichotomous</td>
<td>F_grp: Early/Late (see Tables 3-5, 3-6, 3-7)</td>
</tr>
</tbody>
</table>
Table 3-3. Covariates

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Type</th>
<th>Operational Specification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>Gender.</td>
<td>Dichotomous</td>
<td>Sex: (M/F)</td>
</tr>
<tr>
<td>Race</td>
<td>Race.</td>
<td>Categorical</td>
<td>Race_TXT: White, Non-Hispanic Black, Non-Hispanic Hispanic Asian, Pacific Islander American Indian or Alaskan Unknown / Other</td>
</tr>
<tr>
<td>Income</td>
<td>Median annual income in census tract.</td>
<td>Continuous</td>
<td>($0.00) – from geocoding</td>
</tr>
<tr>
<td>Poverty Index</td>
<td>Percent of children below poverty in census tract.</td>
<td>Continuous</td>
<td>(xx%) – from geocoding</td>
</tr>
<tr>
<td>Prematurity</td>
<td>Gestational age &lt; 37 weeks at birth. (See Appendix B)</td>
<td>Categorical</td>
<td>F_PReM_0 – F_PreM_9 (see Table 3-8) (integer)</td>
</tr>
<tr>
<td>Cesarean Section</td>
<td>Cesarean section as delivery route (763.4)</td>
<td>Dichotomous</td>
<td>F_CS: (integer)</td>
</tr>
<tr>
<td>Low Birth Weight</td>
<td>Low birth weight. (See Appendix C)</td>
<td>Categorical</td>
<td>F_LBW0 - F_LBW9 (See Table 3-9) (integer)</td>
</tr>
<tr>
<td>Extracardiac Anomaly</td>
<td>Additional non-cardiac congenital conditions present at birth. (Any 740.x–744.x, 747.4-9, 748.x-759.x)</td>
<td>Dichotomous</td>
<td>F_EA: (integer)</td>
</tr>
<tr>
<td>Multiple Gestation</td>
<td>Member of multiple gestations. (V31-V37)</td>
<td>Dichotomous</td>
<td>F_MG (integer)</td>
</tr>
<tr>
<td>CHD diagnosis or diagnoses</td>
<td>(Table 3-6)</td>
<td></td>
<td>F_CHD_1 - F_CHD_18 (integer)</td>
</tr>
<tr>
<td>Birth Hospital</td>
<td>Birth Hospital</td>
<td>Categorical</td>
<td>BLNG_Prov_Name</td>
</tr>
<tr>
<td>Months Enrolled</td>
<td>Number of months enrolled from birth through Dec 2011</td>
<td>Integer</td>
<td>Enroll_Mo</td>
</tr>
<tr>
<td>Interhospital Distance</td>
<td>Distance in miles (driving miles) from closest hospital to closest pediatric heart surgery center</td>
<td>Continuous</td>
<td>from geocoding</td>
</tr>
</tbody>
</table>
### Table 3-4. Internal variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Type</th>
<th>Operational Specification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age_last_day</td>
<td>Age at last day of analysis (days)</td>
<td>Continuous</td>
<td>Age (days) on Dec. 31, 2011</td>
</tr>
<tr>
<td>Birthdate</td>
<td>Birth date</td>
<td>Date</td>
<td>bthdate: date at birth YYYY-MM-DD</td>
</tr>
<tr>
<td>Death Date</td>
<td>Death date</td>
<td>Date</td>
<td>d_date: date at death YYYY-MM-DD</td>
</tr>
<tr>
<td>Critical Illness</td>
<td>Diagnosis</td>
<td>Dichotomous</td>
<td>F_CRIT_1 – F_CRIT_14 (see Table 3-7) (integer) 0=N; 1=Y</td>
</tr>
<tr>
<td>Address</td>
<td>Patient Address Of Residence at Birth</td>
<td>String</td>
<td>Internal variable to be used for geocoding – removed after geocoding.</td>
</tr>
</tbody>
</table>

### Table 3-5. Algorithm for assigning early versus late detection

**Group 1. Infants with Late Diagnosis of Critical Congenital Heart Disease**

Conceptual Definition: Neonates presenting with symptoms of critical illness, who have congenital heart disease that was not diagnosed during the birth hospitalization.

Operational Case Definition:
- At least one encounter with diagnosis from Diagnosis List A (congenital heart disease) during days 1-28 of life.
  AND
- Birth hospitalization does NOT have a diagnosis from Diagnosis List A (congenital heart disease).
  AND
- At least one diagnosis from Diagnosis List B (critical illness) at any time during first 28 days of life.

**Group 2. Infants with Early Diagnosis of Critical Congenital Heart Disease**

Conceptual Definition: Neonates with critical congenital heart disease diagnosed during the birth hospitalization.

Operational Case Definition:
- At least one diagnosis from Diagnosis List A (congenital heart disease) made during the birth hospitalization.
### Table 3-6. Diagnosis list A: congenital heart disease

<table>
<thead>
<tr>
<th>ICD-9 code</th>
<th>Diagnosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>745.1x</td>
<td>transposition or double outlet right ventricle</td>
</tr>
<tr>
<td>745.2</td>
<td>Tetralogy of Fallot</td>
</tr>
<tr>
<td>745.3</td>
<td>single ventricle</td>
</tr>
<tr>
<td>745.6</td>
<td>atroventricular canal defect</td>
</tr>
<tr>
<td>746.0x</td>
<td>pulmonary stenosis or atresia</td>
</tr>
<tr>
<td>746.1</td>
<td>tricuspid atresia or stenosis</td>
</tr>
<tr>
<td>746.2</td>
<td>Ebstein’s anomaly</td>
</tr>
<tr>
<td>746.3</td>
<td>aortic stenosis</td>
</tr>
<tr>
<td>746.5</td>
<td>mitral stenosis</td>
</tr>
<tr>
<td>746.7</td>
<td>hypoplastic left heart syndrome</td>
</tr>
<tr>
<td>746.81</td>
<td>subaortic stenosis</td>
</tr>
<tr>
<td>746.83</td>
<td>subpulmonic stenosis</td>
</tr>
<tr>
<td>746.84</td>
<td>other obstructive anomaly or Shone’s complex</td>
</tr>
<tr>
<td>747.1x</td>
<td>coarctation, hypoplasia, interruption of aortic arch</td>
</tr>
<tr>
<td>747.22</td>
<td>aortic atresia</td>
</tr>
<tr>
<td>747.3</td>
<td>pulmonary atresia or hypoplasia</td>
</tr>
</tbody>
</table>

### Table 3-7. Diagnosis list B: critical illness

<table>
<thead>
<tr>
<th>ICD-9 code</th>
<th>Diagnosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>276.2</td>
<td>metabolic and lactic acidosis</td>
</tr>
<tr>
<td>427.5</td>
<td>cardiac or cardiorespiratory arrest</td>
</tr>
<tr>
<td>775.8</td>
<td>acidosis in newborn</td>
</tr>
<tr>
<td>785.51</td>
<td>cardiogenic shock</td>
</tr>
<tr>
<td>785.59</td>
<td>circulatory shock</td>
</tr>
<tr>
<td>785.5</td>
<td>shock unspecified</td>
</tr>
<tr>
<td>785.9</td>
<td>other symptoms involving cardiovascular symptom</td>
</tr>
<tr>
<td>786.03</td>
<td>apnea</td>
</tr>
<tr>
<td>786.06</td>
<td>tachypnea</td>
</tr>
<tr>
<td>786.09</td>
<td>respiratory – other</td>
</tr>
<tr>
<td>798.x</td>
<td>death</td>
</tr>
<tr>
<td>799.0x</td>
<td>asphyxia and hypoxemia</td>
</tr>
<tr>
<td>799.1</td>
<td>respiratory arrest</td>
</tr>
<tr>
<td>799.82</td>
<td>ALTE</td>
</tr>
</tbody>
</table>
Figure 3-2. Hospitals in Texas. Hospitals with pediatric cardiac surgery services in red; all other hospitals in blue. Data from American Hospital Association survey, 2007.
### Table 3-8. Prematurity coding

<table>
<thead>
<tr>
<th>ICD-9 codes</th>
<th>Category Assigned</th>
</tr>
</thead>
<tbody>
<tr>
<td>765.29</td>
<td>Not premature, &gt;37 weeks GA (default)</td>
</tr>
<tr>
<td>765.1, 765.20</td>
<td>Premature, unspecified:</td>
</tr>
<tr>
<td>765.21</td>
<td>Premature, &lt; 24 weeks GA</td>
</tr>
<tr>
<td>765.22</td>
<td>Premature, 24 weeks GA</td>
</tr>
<tr>
<td>765.23</td>
<td>Premature, 25-26 weeks GA</td>
</tr>
<tr>
<td>765.24</td>
<td>Premature, 27-28 weeks GA</td>
</tr>
<tr>
<td>765.25</td>
<td>Premature, 29-30 weeks GA</td>
</tr>
<tr>
<td>765.26</td>
<td>Premature, 31-32 weeks GA</td>
</tr>
<tr>
<td>765.27</td>
<td>Premature, 33-34 weeks GA</td>
</tr>
<tr>
<td>765.28</td>
<td>Premature, 35-36 weeks GA</td>
</tr>
</tbody>
</table>

### Table 3-9. Low birth weight coding

<table>
<thead>
<tr>
<th>ICD-9 codes</th>
<th>Category Assigned</th>
</tr>
</thead>
<tbody>
<tr>
<td>765.09</td>
<td>Normal birthweight &gt;=2499 gm (default)</td>
</tr>
<tr>
<td>V21.30, 764.10, 765.10</td>
<td>LBW, unspecified</td>
</tr>
<tr>
<td>764.01, V21.31</td>
<td>LBW, &lt; 500gm</td>
</tr>
<tr>
<td>764.02, 765.02</td>
<td>LBW, 500-749 gm</td>
</tr>
<tr>
<td>764.03, 765.03, V21.32, 765.00</td>
<td>LBW, 750-999 gm</td>
</tr>
<tr>
<td>764.04, 765.14</td>
<td>LBW, 1000-1249 gm</td>
</tr>
<tr>
<td>764.05, 765.14, V21.33, 765.10</td>
<td>LBW, 1250-1499 gm</td>
</tr>
<tr>
<td>765.06, 765.16</td>
<td>LBW, 1500-1749 gm</td>
</tr>
<tr>
<td>765.07, 765.17, V21.34</td>
<td>LBW, 1750-1999 gm</td>
</tr>
<tr>
<td>765.08, 765.19</td>
<td>LBW, 2000-2499 gm</td>
</tr>
</tbody>
</table>
CHAPTER 4
RESULTS

Sample Size Determination

Preliminary data consisted of subjects with CCHD (n=1711) born in 2011 with cost follow-up data through December 1, 2011. Based on analysis of the preliminary data, we calculated clustering by birth hospital with 12 members per group (in this case obtained by taking the harmonic mean of the cluster sizes in the preliminary dataset), an ICC of 0.16, an early detection group 3 times the size of the late detection group, a mean log-transformed cost difference detection threshold of 0.6 between early and late groups, and a standard deviation of 1.78. Based upon these assumptions, a power of 0.9 with α=0.05, to detect half of the observed difference in cost with twice the observed variance, would require a sample size of 940 subjects. Further uncertainty was introduced by the challenge of unbalanced cluster sizes, so our goal for final sample size was to have at least twice the calculated size, or 1880 subjects.

Preliminary Analysis

The final dataset contained data from 3267 subjects with CCHD, 2602 (79.6%) of whom had early detection of CCHD and 665 (20.4%) of whom had late detection. The overall rate of CCHD based upon 806,841 live births in Texas in 2008 and 2009 was 4 per 1000. This is consistent with previously published estimates.\textsuperscript{1}

Missingness

Missingness is summarized in Table 4-1.

The primary outcome variable was missing in only one case, but it was zero in 69 additional cases that were therefore unsuitable for subsequent logarithmic
transformations. These observations were assumed to be erroneous, as a cost of $0 for a child with CCHD is implausible, and they were dropped from all outcome analyses.

The CRG was missing in 414 (12.7%) cases, 326 (12.5%) in the early detection group and 88 (13.2%) in the late detection group. The proportion missing was not significantly different between groups (p=0.6, chi-square).

The interhospital distance was missing in 104 cases (3.2%), 91 (3.5%) in the early detection group and 13 (2.3%) in the late detection group. These data were missing in subjects for whom addresses or zip codes in Texas were not available from the original claims data. The proportion missing was not significantly different between groups (p=0.1, chi-square). These cases were dropped from all instrumental variable and treatment effect models but analyzed in the final models.

Income and poverty data were missing in 443 (13.6%) subjects, 352 (13.5%) in the early detection group and 91 (13.7%) in the late detection group. These data were missing in those who not have street addresses that mapped appropriately to the census tracts used for geocoding. The proportion missing was not significantly different between groups (p=0.9, chi-square).

Gender was missing for one subject. All other variables were extracted and coded with a default value and thus had no missing values.

**Normality Testing**

The distribution of raw cost is shown in Figure 4-1. Based on unacceptably high skewness (5.1) and kurtosis (39.5) it was felt to be unsuitable for models that assume a normal distribution of the outcome variable. The distribution of the natural logarithm of the raw cost (log-transformed cost) is shown in Figure 4-2. It had very low skewness (-
0.2) and kurtosis (-0.01). Outcome modeling used the log-transformed cost or the raw cost with a log link function, depending upon the specific model.

**Descriptive Statistics**

The baseline characteristics of all variables and differences between early and late detection groups are given in Table 4-2 for continuous variables and in Table 4-3 for categorical variables. CRG was analyzed as a continuous and a categorical variable for this purpose, but is omitted from Table 4-3 for simplicity. Cost, log-transformed cost (see Figure 4-3), CRG (see Figure 4-4), interhospital distance (see Figure 4-5), gender, low birth weight as a binary variable, the presence of multiple gestation, the presence of extracardiac anomaly, and aortic arch obstruction all showed a significant difference between detection groups. Median income by census tract, percent of children in poverty by census tract, enrolled months, race, prematurity as a binary variable, and single ventricle physiology did not show a significant difference between detection groups.

The rates of prematurity and low birth weight were 51.5% and 67.8% in the overall study sample. The baseline rate of prematurity in the US is one in nine (11.1%)\(^{117}\)

**Intra-class Correlation**

There were 177 (5.4%) of subjects with no birth hospital available. When ignoring these observations, the ICC was 0.055. When counting these observations as clustered around a single subject, the ICC was 0.048.

**Sample Size Recalculation**

When compared with the preliminary dataset, the final dataset had a much lower intra-class correlation coefficient (0.048-0.055), a different ratio of early to late detection
(4:1), a different average cluster size (harmonic mean 2.3), and a smaller difference between log-transformed cost between the early and late detection groups (0.42), an additional post-hoc sample size calculation was performed based upon the actual characteristics of the final dataset. When assuming the “worst-case” scenario, with a harmonic mean cluster size of 2, an ICC of 0.57, power of 0.9 to detect a difference of one half of the actual difference observed, with an α of 0.05, required 2280 total subjects.

**Heteroskedasticity**

The Park test based on a first-stage OLS outcomes model yielded a $\lambda_1$ of 1.9. The regression is shown in Figure 4-6. The Park tests using a first-stage GLM with Poisson and gamma distributions yielded $\lambda_1$ values of 1.6 and 1.8 respectively. The poverty and income variables were left out of the first-stage models due to missingness; all other covariates were included.

Based on the consistent result of the Park tests in characterizing the heteroskedasticity present in the predicted outcome variable, we chose to use a gamma distribution for our final GLM approach.

**Selection Bias**

**Instrumental Variable Analysis**

Using the first stage of instrumental variable, the model given in Equation (3-6) was run using four sets of covariates: all covariates, all covariates except the highly missing income and poverty variables, a minimal model consisting of non-white race, prematurity, low birth weight, extracardiac malformation, single ventricle physiology, and arch obstruction, and a model with no covariates. In each case, the instrumental variable $z$ was found to be only weakly, although significantly, related to the group
selection variable. Parameter estimates for \( \pi_1 \) ranged from -0.004 and -0.00408. Using baseline probability of early detection of 0.8, a change in the interhospital distance by one standard deviation, or 45 minutes, the change in probability of being in the early detection group would be given by

\[
\pi_1(p)(1-p)\Delta z = 0.004(0.8)(0.2)(45) = 0.028,
\]

or a 2.8% absolute change in probability

Despite the weak instrument, we prepared the second stage of the three-stage instrumental variable analysis as planned. After modeling the second stage given in Equation (3-7), the predictor variable \( \phi \), which incorporated the effect of regressors included in Equation (3-6) was now strongly related to the selection variable.

**Heckman Selection Model**

The Heckman selection model yielded a \( \rho \) statistic of 0.16 (95% CI -1.0-0.40). This was not significantly different than 0 (\( p=0.32 \)), and thus we found no evidence of group selection bias via endogeneity of the selection variable.

**Measured Variables That Influenced Group Selection**

In both the instrumental variable modeling and the treatment effects model, variables that consistently and significantly predicted a higher probability of early detection were female gender, multiple gestation, delivery via Cesarean section, and univentricular physiology. Variables that significantly predicted a lower probability of early detection were the presence of an extracardiac lesion and aortic arch obstruction.

**Modeling the Primary Hypothesis**

Using a GLM with a log link, a gamma distribution, an exchangeable covariance matrix, and clustering by birth hospital, we first modeled the outcome equation using all covariates, including the months enrolled. This model had an high level of missing data,
with 501 (15.3%) of observations dropped, primarily due to the SES variables. Based on the results of this model as well as the preliminary OLS model used in the Park test, we selected significant and near-significant covariates to include in the parsimonious model. This model represented our best effort at incorporating all preliminary analyses and testing into a single model with minimal missingness. Only 70 observations (2.1%) were dropped from this final analysis, all due to missing or zero cost data.

We then repeated the parsimonious model without accounting for clustering effects, and also compared the results to the OLS model using the same parameters that was previously run as part of the Park test. Because of a weak instrument and no evidence of selection bias based on the Heckman model, we did not include the predicted selection probability from the instrumental variable approach in the final modeling process.

In the “best” model, early detection was associated with a decreased cost (p<0.009). The parameter estimate for this effect was -0.18, which when exponentiated yields 0.84, or a 16% reduction in base cost. This would give a cost savings of approximately $18,000 compared to late detection, using the observed mean cost as a baseline. Significant predictors of increased cost were non-white race ($34,000 cost increase), the presence of an extracardiac lesion ($223,000), single ventricle physiology ($135,000), and aortic arch obstruction ($56,000). Other significant predictors of decreased cost were prematurity ($37,000 cost reduction) and low birth weight ($44,000 cost reduction), although concerns about the robustness of the prematurity and low birth weight are discussed below. The log-transformed months enrolled did not significantly
predict cost variation. The results of the “best” model as well as the three comparison models are presented together in Table 4-4.

For the sake of transparency, Appendix A also reports the general results of additional exploratory models that were run with instrumental variables, alternative distributions, and/or an offset variable. Each of these models was not used to inform the final analysis of the primary hypothesis for reasons given above. Appendix B gives the SAS and STATA code for all data manipulation and analyses performed during this study.

**Testing Secondary Aims**

**Racial/Ethnic and SES Disparity**

We conducted analysis of racial and ethnic disparities across race/ethnicity and SES for the outcome variables, the selection variable, and all covariates.

**Outcome variables**

Raw and log-transformed cost both differed across racial/ethnic categories (p<0.0001, ANOVA). The log-transformed cost by racial/ethnic category is shown in Figure 4-7. Although the raw cost was higher in nonwhite vs. white subjects ($117,800 versus $93,500, p=0.04, t-test), the log-transformed cost was not significantly different using this simplified grouping. In the outcomes model used for the primary hypothesis, non-white race/ethnicity was positively associated with increased cost using the simplified binary classification, but only “Unknown / Other” was a consistent predictor when all categories were analyzed.

The SES variables, median income and the percent of children living in poverty, were not significant predictors of cost.
The CRG score as a continuous variable was not significantly different across racial categories or when comparing white and non-white subjects. There was, however, a difference across racial categories when considering CRG as a categorical variable. In this analysis 42% of subjects with "Unknown / Other" race/ethnicity were in CRG category 5, "Special Health Care Needs – Major," while only 25.5% of overall subjects were in this category.

**Selection variable**

The proportion of subjects in each racial/ethnic category did not differ significantly between the early and late detection groups, nor did the proportion of white vs. non-white subjects differ between the groups (see Table 4-3). Similarly, race/ethnicity, whether considered by category or as a binary white/non-white variable, was not significantly associated with the selection variable in the instrumental variable or treatment effects model.

Income and poverty did not significantly predict the selection variable in the maximally specified instrumental variable model.

**Covariates**

There was no significant difference in gender across racial/ethnic categories. Prematurity was different across racial categories (p<0.0001) and appeared lower in "Black, Non-Hispanic" subjects (37.3%) than in all others (47.3-55.2%). Low birth weight was similarly different across racial/ethnic categories (p=0.0005) and appeared lower in "Black, Non-Hispanic" and "Unknown/Other" subjects (59.9-63.4%) than in all others (69.8-78.3%). Multiple gestation was also significantly different across racial/ethnic categories, occurring more frequently in "Black, Non-Hispanic" subjects (16.5%) than in all others (7.7-10.2%). Note that all pairwise comparisons are post-hoc
and were not analyzed for statistical significance. There was no significant difference in Cesarean section rates across racial/ethnic groups. The proportion of those with aortic arch obstruction and single ventricle physiology differed among groups but with no discernible pattern.

When using the binary white/non-white binary race/ethnicity variable, no categorical covariates were significant; however, nonwhite versus white subjects had a lower mean median income for census tract ($37,900 versus $49,000, p < 0.0001, see Figure 4-8) and a higher percent of children living in poverty for census tract (26.1% versus 15.6%, p < 0.0001, see Figure 4-9).

Finally, the interhospital distance was significantly different across racial/ethnic categories (p<0.0001) by ANOVA and significantly shorter in nonwhite versus white subjects (36.4 versus 46.3 minutes, p<0.0001).

SES variables were not tested directly against additional covariates beyond what was described above.

**Overall Disease Burden**

The mean CRG was slightly higher in the late versus early detection group (3.4 versus 3.2, p<0.002 by t-test). When considering the CRG as a categorical variable, the difference in distributions between detection groups was significantly different.(p=0.001 by chi-square). Of those with an available CRG score, 31.7% of the late detection group versus 23.9% of the early detection group fell into CRG category 5, “Special Health Care Needs – Major.” Similarly, 18.4% of the late detection group versus 21.4% of the early detection group, fell into CRG category 1, “Healthy.” Due to the high rate of missingness of the CRG as well as uncertainty regarding its timing, regression analyses were not performed for this variable.
Survival

Only 10 (0.3%) subjects were coded as non-survivors using the algorithm given in Table 3-1. This method, using the 798.x group of ICD-9 codes, identified far fewer deaths than would be expected based on well-established survival rates.\(^5\) Our assumption is that death was inconsistently recorded as a diagnosis code, and we were thus unable conduct this analysis using the available claims data.
Table 4-1. Missing data

<table>
<thead>
<tr>
<th>Variable</th>
<th>Total No.</th>
<th>Missing No.</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost</td>
<td>3266</td>
<td>1</td>
<td>0.0</td>
</tr>
<tr>
<td>Log-Transformed Cost</td>
<td>3197</td>
<td>70</td>
<td>2.1</td>
</tr>
<tr>
<td>Interhospital Distance</td>
<td>3163</td>
<td>104</td>
<td>3.2</td>
</tr>
<tr>
<td>Income</td>
<td>2824</td>
<td>443</td>
<td>13.6</td>
</tr>
<tr>
<td>Poverty</td>
<td>2824</td>
<td>443</td>
<td>13.6</td>
</tr>
<tr>
<td>Enrolled Months</td>
<td>3267</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>CRG</td>
<td>2853</td>
<td>414</td>
<td>12.7</td>
</tr>
<tr>
<td>Gender</td>
<td>3266</td>
<td>1</td>
<td>0.0</td>
</tr>
<tr>
<td>Race/Ethnicity</td>
<td>3267</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>Gestational Age</td>
<td>3267</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>Birth Weight</td>
<td>3267</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>Multiple Gestation</td>
<td>3267</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>Cesarean</td>
<td>3267</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>Extracardiac Lesion</td>
<td>3267</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>Arch Obstruction</td>
<td>3267</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>Coarctation</td>
<td>3267</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>Single Ventricle</td>
<td>3267</td>
<td>0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

Table 4-2. Descriptive statistics for continuous data

<table>
<thead>
<tr>
<th>Variable</th>
<th>Overall</th>
<th>Early Detect.</th>
<th>Late Detect.</th>
<th>p&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>Mean</td>
</tr>
<tr>
<td>Cost&lt;sup&gt;a&lt;/sup&gt;</td>
<td>114.7</td>
<td>233.8</td>
<td>108.3</td>
<td>138.4</td>
</tr>
<tr>
<td>Log-Transformed Cost</td>
<td>10.4</td>
<td>1.7</td>
<td>10.3</td>
<td>10.8</td>
</tr>
<tr>
<td>Interhospital Dist. (min.)</td>
<td>37.7</td>
<td>45.4</td>
<td>35.9</td>
<td>44.9</td>
</tr>
<tr>
<td>Income&lt;sup&gt;a&lt;/sup&gt;</td>
<td>40.3</td>
<td>18.3</td>
<td>40.1</td>
<td>41.2</td>
</tr>
<tr>
<td>Poverty (%)</td>
<td>24.6</td>
<td>14.1</td>
<td>24.7</td>
<td>24.0</td>
</tr>
<tr>
<td>Enrolled Months (mo.)</td>
<td>27.2</td>
<td>12.3</td>
<td>27.1</td>
<td>27.4</td>
</tr>
<tr>
<td>CRG</td>
<td>3.3</td>
<td>1.5</td>
<td>3.2</td>
<td>3.4</td>
</tr>
</tbody>
</table>

<sup>a</sup>thousands of U.S. Dollars

<sup>b</sup>by t-test with equal or unequal variances as indicated by Folded F test
Table 4-3. Descriptive characteristics of categorical data

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Total N (%)</th>
<th>Early Detect. N (% of group)</th>
<th>Late Detect. N (% of group)</th>
<th>p&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>1758 (53.8)</td>
<td>1364 (52.4)</td>
<td>394 (59.3)</td>
<td>0.0014</td>
</tr>
<tr>
<td>Female</td>
<td>1508 (46.2)</td>
<td>1238 (47.6)</td>
<td>270 (40.7)</td>
<td></td>
</tr>
<tr>
<td>White/Non-Hispanic</td>
<td>450 (13.8)</td>
<td>353 (13.6)</td>
<td>97 (14.6)</td>
<td>0.69</td>
</tr>
<tr>
<td>Black/Non-Hispanic</td>
<td>431 (13.2)</td>
<td>299 (11.5)</td>
<td>83 (12.5)</td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>1972 (60.4)</td>
<td>1581 (60.8)</td>
<td>391 (58.8)</td>
<td></td>
</tr>
<tr>
<td>Asian/Pacific Isl.</td>
<td>23 (0.7)</td>
<td>17 (0.7)</td>
<td>6 (0.9)</td>
<td></td>
</tr>
<tr>
<td>Amer. Indian/Alaskan</td>
<td>6 (0.2)</td>
<td>4 (0.2)</td>
<td>2 (0.3)</td>
<td></td>
</tr>
<tr>
<td>Unknown/Other</td>
<td>385 (11.8)</td>
<td>299 (11.5)</td>
<td>86 (12.9)</td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>450 (13.8)</td>
<td>353 (13.6)</td>
<td>97 (14.6)</td>
<td>0.50</td>
</tr>
<tr>
<td>Non-White</td>
<td>2817 (86.2)</td>
<td>2249 (86.4)</td>
<td>568 (85.4)</td>
<td></td>
</tr>
<tr>
<td>Term (≥ 37 weeks)</td>
<td>1584 (48.5)</td>
<td>1249 (48.0)</td>
<td>335 (50.4)</td>
<td>0.27</td>
</tr>
<tr>
<td>Preterm (&lt;37 weeks)</td>
<td>1683 (51.5)</td>
<td>1353 (52.0)</td>
<td>330 (49.6)</td>
<td></td>
</tr>
<tr>
<td>Birth Weight ≥ 2.5 kg</td>
<td>1052 (32.2)</td>
<td>793 (30.5)</td>
<td>259 (39.0)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Birth Weight &lt; 2.5 kg</td>
<td>2215 (67.8)</td>
<td>1809 (69.5)</td>
<td>406 (61.1)</td>
<td></td>
</tr>
<tr>
<td>Singleton</td>
<td>2956 (90.5)</td>
<td>2330 (89.5)</td>
<td>626 (94.1)</td>
<td>0.0003</td>
</tr>
<tr>
<td>Multiple gestation</td>
<td>311 (9.5)</td>
<td>272 (10.5)</td>
<td>39 (5.9)</td>
<td></td>
</tr>
<tr>
<td>Vaginal Delivery</td>
<td>2820 (86.3)</td>
<td>2224 (85.5)</td>
<td>596 (89.6)</td>
<td>0.005</td>
</tr>
<tr>
<td>Cesarean Delivery</td>
<td>447 (13.7)</td>
<td>378 (14.5)</td>
<td>69 (10.4)</td>
<td></td>
</tr>
<tr>
<td>No Extracardiac Anomaly</td>
<td>1590 (48.7)</td>
<td>268 (40.3)</td>
<td>1322 (50.8)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Extracardiac Anomaly</td>
<td>1677 (51.3)</td>
<td>397 (59.7)</td>
<td>1280 (49.2)</td>
<td></td>
</tr>
<tr>
<td>Normal Arch</td>
<td>2530 (77.4)</td>
<td>2037 (78.3)</td>
<td>493 (74.4)</td>
<td>0.02</td>
</tr>
<tr>
<td>Arch Obstruction</td>
<td>737 (22.6)</td>
<td>565 (21.7)</td>
<td>172 (25.9)</td>
<td></td>
</tr>
<tr>
<td>Two Ventricle</td>
<td>2274 (84.9)</td>
<td>2203 (84.7)</td>
<td>571 (85.9)</td>
<td>0.44</td>
</tr>
<tr>
<td>Single Ventricle</td>
<td>493 (15.1)</td>
<td>399 (15.3)</td>
<td>94 (14.1)</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>greater than chi-square
Figure 4-1. Distribution of raw cost

Figure 4-2. Distribution of log-transformed cost
Figure 4-3. Comparison of log-transformed cost (0=late detection, 1=early detection)

Figure 4-4. Comparison of CRG (0=late detection, 1=early detection)
Figure 4-5. Comparison of interhospital distance (0=late detection, 1=early detection)
Figure 4-6. Park test for heteroskedasticity (using OLS as first-stage model)
### Table 4-4. Final outcome model results

<table>
<thead>
<tr>
<th>Variable</th>
<th>Parameter</th>
<th>&quot;Best&quot; Model&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Maximal&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Non-Multilevel&lt;sup&gt;c&lt;/sup&gt;</th>
<th>OLS&lt;sup&gt;d&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>( \beta_0 )</td>
<td>10.9</td>
<td>10.8</td>
<td>10.9</td>
<td>9.6</td>
</tr>
<tr>
<td>Early Detection</td>
<td>( \beta_1 )</td>
<td>-0.17&lt;sup&gt;*&lt;/sup&gt;</td>
<td>-0.13</td>
<td>-0.16&lt;sup&gt;*&lt;/sup&gt;</td>
<td>-0.22&lt;sup&gt;*&lt;/sup&gt;</td>
</tr>
<tr>
<td>Female</td>
<td>( \beta_1 )</td>
<td>-</td>
<td>0.003</td>
<td>-</td>
<td>-0.09</td>
</tr>
<tr>
<td>Premature</td>
<td>( \beta_1 )</td>
<td>-0.39&lt;sup&gt;*&lt;/sup&gt;</td>
<td>-0.39</td>
<td>-0.48&lt;sup&gt;*&lt;/sup&gt;</td>
<td>-0.63&lt;sup&gt;*&lt;/sup&gt;</td>
</tr>
<tr>
<td>Low Birth Weight</td>
<td>( \beta_1 )</td>
<td>-0.49&lt;sup&gt;*&lt;/sup&gt;</td>
<td>-0.47</td>
<td>-0.47&lt;sup&gt;*&lt;/sup&gt;</td>
<td>-0.48&lt;sup&gt;*&lt;/sup&gt;</td>
</tr>
<tr>
<td>Multiple Gestation</td>
<td>( \beta_1 )</td>
<td>-</td>
<td>-0.04</td>
<td>-</td>
<td>-0.04</td>
</tr>
<tr>
<td>Cesarean Section</td>
<td>( \beta_1 )</td>
<td>-</td>
<td>-0.10</td>
<td>-</td>
<td>-0.10</td>
</tr>
<tr>
<td>Extracardiac Lesion</td>
<td>( \beta_1 )</td>
<td>1.08&lt;sup&gt;*&lt;/sup&gt;</td>
<td>1.05&lt;sup&gt;*&lt;/sup&gt;</td>
<td>1.09&lt;sup&gt;*&lt;/sup&gt;</td>
<td>1.06&lt;sup&gt;*&lt;/sup&gt;</td>
</tr>
<tr>
<td>Single Ventricle</td>
<td>( \beta_1 )</td>
<td>0.78&lt;sup&gt;*&lt;/sup&gt;</td>
<td>0.78&lt;sup&gt;*&lt;/sup&gt;</td>
<td>0.78&lt;sup&gt;*&lt;/sup&gt;</td>
<td>1.02&lt;sup&gt;*&lt;/sup&gt;</td>
</tr>
<tr>
<td>Aortic Arch Obstruction</td>
<td>( \beta_1 )</td>
<td>0.40&lt;sup&gt;*&lt;/sup&gt;</td>
<td>0.41&lt;sup&gt;*&lt;/sup&gt;</td>
<td>0.52&lt;sup&gt;*&lt;/sup&gt;</td>
<td>0.47&lt;sup&gt;*&lt;/sup&gt;</td>
</tr>
<tr>
<td>Median Income&lt;sup&gt;e&lt;/sup&gt;</td>
<td>( \beta_1 )</td>
<td>-</td>
<td>0.0</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>% Children in Poverty&lt;sup&gt;e&lt;/sup&gt;</td>
<td>( \beta_1 )</td>
<td>-</td>
<td>-0.0004</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Black, Non-Hispanic&lt;sup&gt;f&lt;/sup&gt;</td>
<td>( \beta_1 )</td>
<td>-</td>
<td>0.24&lt;sup&gt;*&lt;/sup&gt;</td>
<td>-</td>
<td>0.004</td>
</tr>
<tr>
<td>Hispanic&lt;sup&gt;f&lt;/sup&gt;</td>
<td>( \beta_1 )</td>
<td>-</td>
<td>0.16</td>
<td>-</td>
<td>-0.12</td>
</tr>
<tr>
<td>Amer. Indian / Alaskan Native&lt;sup&gt;f&lt;/sup&gt;</td>
<td>( \beta_1 )</td>
<td>-</td>
<td>0.47&lt;sup&gt;*&lt;/sup&gt;</td>
<td>-</td>
<td>1.07</td>
</tr>
<tr>
<td>Asian/Pacific Islander&lt;sup&gt;f&lt;/sup&gt;</td>
<td>( \beta_1 )</td>
<td>-</td>
<td>0.40</td>
<td>-</td>
<td>0.25</td>
</tr>
<tr>
<td>Other /Unknown&lt;sup&gt;f&lt;/sup&gt;</td>
<td>( \beta_1 )</td>
<td>-</td>
<td>0.52&lt;sup&gt;*&lt;/sup&gt;</td>
<td>-</td>
<td>0.20&lt;sup&gt;*&lt;/sup&gt;</td>
</tr>
<tr>
<td>All Non-White&lt;sup&gt;f&lt;/sup&gt;</td>
<td>( \beta_1 )</td>
<td>0.26&lt;sup&gt;*&lt;/sup&gt;</td>
<td>-</td>
<td>0.13&lt;sup&gt;*&lt;/sup&gt;</td>
<td>-</td>
</tr>
<tr>
<td>Log of Months Enrolled</td>
<td>( \beta_1 )</td>
<td>0.04</td>
<td>0.04</td>
<td>0.01</td>
<td>0.28&lt;sup&gt;*&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

<sup>a</sup>Parsimonious GLM with significant covariates, multilevel model
<sup>b</sup>Maximal GLM with all covariates, multilevel model
<sup>c</sup>Parsimonious GLM with significant covariates, no multilevel model
<sup>d</sup>OLS with all covariates except SES, no multilevel model
<sup>e</sup>by Census Tract
<sup>f</sup>with "White" race/ethnicity as referent group
<sup>*</sup>p < 0.05
<sup>-</sup>- indicates not modeled
Figure 4-7. Log-transformed cost by racial/ethnic category
Figure 4-8. Median income by racial/ethnic category

Figure 4-9. Percent of children in poverty in census tract by race/ethnicity
CHAPTER 5
DISCUSSION

Primary Hypothesis

Our results support our primary hypothesis, demonstrating that early detection of CCHD predicts an average reduction in long-term health-care cost of approximately 16%. In our study this translated to approximately $18,300 in unadjusted U.S. dollars. This could partially or completely offset the incremental cost-effectiveness ratio of an NBS program.

We reached this conclusion using a modeling strategy that quantitatively evaluated (1) missingness, (2) distribution of the outcome variable, (3) clustering and intra-class correlation, (4) heteroskedasticity, and (5) endogeneity and group selection bias. In addition, our primary result reached significance in two comparison models using different modeling methodologies, and reached near-significance in a model that differed only by the incorporation of additional covariates.

We also found that the presence of an extracardiac lesion, single ventricle physiology, aortic arch obstruction, and non-white race were strong predictors of increased cost across models, with particularly strong effects from extracardiac lesions and single ventricle physiology. (Hypothesis 2)

Prematurity and low birth weight predicted decreased cost. Given the significantly higher mortality of premature infants with CCHD, in particular those weighing less than 1500 grams at birth, there may have been earlier mortality and thus decreased long-term cost in this group. However, there was also an unexpectedly high prevalence of prematurity (51.5%) and low birth weight (67.8%) in our study sample. There is a higher frequency of CCHD in very low birth weight
infants\textsuperscript{89} and a clinical practice of preterm delivery or induction persists despite strong evidence against it,\textsuperscript{118} so some degree of overrepresentation would be expected in this population. However, the rates of both prematurity and low birth weight in our study were approximately five times baseline population rates.\textsuperscript{117} There may be a flaw in recording the diagnoses of prematurity and low birth weight in claims data and/or a flaw in our algorithm for converting it to a binary variable. For this reason we urge caution when interpreting these particular relationships.

**Secondary Aims**

**Racial/Ethnic and SES Disparities**

Non-white race was a predictor of increased cost in the final outcomes model. (Hypothesis 3A). The reason for this is unclear.

However, while many covariates including prematurity, multiple gestation, Cesarean section, aortic arch obstruction, and single ventricle physiology were different across racial/ethnic categories, these differences were not significant across a binary white/non-white comparison. This would suggest that a binary comparison does not capture the texture of racial variation and that the minimal model was perhaps too reductive. This is further reinforced by the difficulty in identifying a clear qualitative pattern across the full range or racial/ethnic categories for multiple variables, and the large “Unknown/Other” category. For future studies a more robust system of racial/ethnic classification would be preferred if available.

Race/ethnicity did not appear to predict early versus late detection of CCHD.

SES variables did not appear to predict either the selection or primary outcome, cost, although the degree of missingness limited their utility significantly and they were not included in the final modeling strategy. (Hypothesis 3B)
Finally, our results confirmed the presence of baseline economic disparities across racial/ethnic groups, demonstrating lower median incomes and a higher percent of children living in poverty near the homes of Black and Hispanic subjects in particular. The interhospital distance varied by race/ethnicity, suggesting that subjects in some categories tend to live closer to pediatric cardiac surgery centers.

**Overall Disease Burden**

The significantly higher mean CRG score in the late detection group, the higher percentage of late versus early detection subjects in the most severe CRG category, and the lower percentage of late versus early detection subjects in the least severe CRG category all suggest that late detection of CCHD is associated with an increased long-term disease burden affecting resource utilization. Because the pairwise percentage comparisons were post-hoc and quantitative, we regard this as evidence to support but not prove Hypothesis 4.

**Limitations and Threats to Validity**

Multiple threats to validity affect any quasi-experimental study. We follow Shadish, Cook, and Campbell’s excellent text in explicating potential threats to validity and how they affected or were addressed in this study.\(^{119}\)

**Threats to Statistical Conclusion Validity**

We used a prospective power calculation and verified it with a post-hoc calculation to avoid *low statistical power*. To avoid violating specific *statistical assumptions*, we used methods to test and account for distributional assumptions, heteroskedasticity, and clustering. To *avoid inflating the Type I error rate*, we accounted for clustered data and avoided multiple comparisons except where clearly stated. *Unreliability of measures* was an issue with geocoded data, survival,
prematurity and low birth weight diagnoses, and a small proportion of the cost variables. 

*Heterogeneity of units* was also suggested by differences in baseline data among groups, and was addressed by controlling for covariates in the final outcomes model. Threats related to *treatment implementation unreliability* and *experimental setting variability* did not apply to this quasi-experimental and retrospective study.

**Threats to Internal Validity**

Selection was the biggest source of potential bias in this study. *Sample selection* was an issue both in limiting the study to one state and limiting it to Medicaid and CHIP clients. This may have skewed the data towards groups who differ in characteristics influencing the outcome. Race/ethnicity and SES are two means by which this could occur, both of which are analyzed in some detail in the text. Future studies can avoid this issue by including geographically and demographically diverse states as well as using all-payer claims databases. *Group selection* has already been addressed in detail.

*Ambiguous temporal precedence* was not an issue as all explanatory variables were present at birth and all outcome variables developed after birth. *History* may have been an issue in this particular birth cohort from 2008-2009 although preliminary results from infants born in 2011 are very similar. More generally, continuous advances and changing technology in the diagnosis and treatment of CHD make it difficult, as in any field, to compare estimates of cost, survival, or comorbidity across time. *Maturation* is not a major issue as our subjects are followed from birth on. Similarly, *regression to the mean* does not apply as our subjects are selected based upon a relatively fixed diagnosis. *Attrition* due to disenrollment in Medicaid/CHIP is a potential threat to validity here, but inclusion of a term for total months enrolled in the final outcome model
controlled for this issue. **Testing** is not an issue although in this study although would be for future prospective studies involving long-term neurocognitive outcomes.

**Instrumentation**, in the form of changing trends in recording ICD-9 codes, may affect results although this is unlikely in such a short period of time. In the future, however, the use of ICD-10 codes may make comparison with ICD-9 based data very challenging. Finally, specific threats related to groups undergoing treatments, including **treatment diffusion, compensatory equalization, compensatory rivalry, and resentful demoralization**, do not apply to this study.

**Threats to Construct Validity**

There are several potential and important threats to construct validity in this study. **Inadequate explication of constructs** may affect our definition and operationalization of variables; indeed, this appeared to be the case with the prematurity and low birth weight variables. We guarded against this by carefully defining variable and data specifications as well as carefully examining a preliminary data set, making specification adjustments, and re-examining the final dataset. **Mono-operation bias** is unlikely as children were treated at multiple hospitals. Although we identified several outcome measures in part to avoid **mono-method bias**, we completed the full, planned analysis only on cost, so this remains a threat to validity.

Threats not applicable to this study methodology, which did not have treatment, study participants per se, or any staff interactions with subjects, included **hypothesis guessing, evaluation apprehension, experimenter expectancies, novelty or disruption effects, interaction of testing or treatment effects, confounding constructs with levels of constructs, interaction of testing and treatment, and interaction of different treatments.**
Threats to External Validity (Generalizability)

The primary threats to the external validity of this study is the interaction of selection with the causal relationship. Selecting based upon enrollment in Medicaid and CHIP makes it imperative to repeat this type of study with an all-payer claims database, as well as with subjects whose racial/ethnic and economic diversity reflects the makeup of the population to which we wish to generalize. In a more limited sense, interaction of the setting—Texas—with the causal relationship should be addressed in similar fashion in the future by studying other states or regions. Future studies could also include non-US locations with qualitatively different health-care systems, to avoid context-dependent mediation. The interaction of the causal relationship with treatment variations could become an issue in studies attempting to manipulate the “treatment” variable, here the diagnosis of CCHD. Finally, interaction of the causal relationship with units is less of an issue with this study as individual people are the sampling unit of interest.

Additional study limitations

We did not adjust cost values for inflation by year. Although the time span of this study was short enough that it may not have affected, future and more comprehensive studies should adopt this methodology, particularly if researchers hope to inform or affect policy. The study also would have been significantly strengthened by adopting a more robust method to evaluate survival, such as a link with a publicly available death registry. Similarly, obtaining hospital records for a sample of patients to verify the diagnoses and other information would have helped assure the quality of the claims database variable extraction process.
Future Directions

Given the legislative landscape in the United States, the time is ripe for prospective experimental and/or quasi-experimental studies on a larger scale to determine the potential cost benefit realized by increasing the early detection of congenital heart disease, as well as the potential benefit to long-term chronic disease burden, which in turn affects resource utilization. Long-term neurocognitive outcomes are also important areas of study and could potentially impact health care cost and resource utilization in this important population. We envision a multi-state study of a newborn screening program, combining all-payer claims data over time with a population subsample undergoing periodic neurocognitive screening and reporting on health related quality of life using an instrument designed to meet the needs of children and young adults with CCHD.120-122 This analysis would be compared to actual economic data on the cost of implementation of the screening program, including an incremental cost-effectiveness ratio based on measured, rather than forecast, costs.

Conclusion

Diagnosis of CCHD before initial hospital discharge, when compared with later diagnosis, predicted a reduction in long-term health care costs in this sample of subjects with CCHD enrolled in Texas Medicaid and CHIP and born in 2008-2009. Non-white race, extracardiac lesions, aortic arch obstruction, and single ventricle physiology all predicted significantly increased cost. In addition, the long-term disease burden, which potentially affects resource utilization, appears to be less in the group with early than with late diagnosis. Prospective studies with more diverse populations, payer mixes, and geographic regions are needed to confirm these findings.
<table>
<thead>
<tr>
<th>Multi-level</th>
<th>Instrument</th>
<th>Co-variates</th>
<th>Offset</th>
<th>Distr.</th>
<th>Cov. Struct.</th>
<th>$\beta_1$</th>
<th>$p$</th>
<th>Significant Covariates</th>
</tr>
</thead>
<tbody>
<tr>
<td>yes</td>
<td>2 stage</td>
<td>maximal</td>
<td>no</td>
<td>gamma</td>
<td>exch.</td>
<td>-0.44</td>
<td>0.65</td>
<td>premature, lowbirthwt, extracardiac, univentricular, archobstruction, black or Am. Indian race</td>
</tr>
<tr>
<td>yes</td>
<td>3 stage</td>
<td>maximal</td>
<td>no</td>
<td>gamma</td>
<td>exch.</td>
<td>-0.15</td>
<td>0.85</td>
<td>premature, lowbirthwt, extracardiac, univentricular, archobstruction, black, Am. Indian, unknown race</td>
</tr>
<tr>
<td>no</td>
<td>2 stage</td>
<td>minimal</td>
<td>no</td>
<td>gamma</td>
<td>exch.</td>
<td>-0.28</td>
<td>0.48</td>
<td>premature, lowbirthwt, extracardiac, univentricular, archobstruction, nonwhite</td>
</tr>
<tr>
<td>no</td>
<td>3 stage</td>
<td>minimal</td>
<td>no</td>
<td>gamma</td>
<td>exch.</td>
<td>-0.44</td>
<td>0.27</td>
<td>premature, lowbirthwt, extracardiac, univentricular, archobstruction, nonwhite</td>
</tr>
<tr>
<td>yes</td>
<td>2 stage</td>
<td>minimal</td>
<td>no</td>
<td>gamma</td>
<td>exch.</td>
<td>-0.92</td>
<td>0.1</td>
<td>premature, lowbirthwt, extracardiac, univentricular, archobstruction, nonwhite</td>
</tr>
<tr>
<td>yes</td>
<td>3 stage</td>
<td>minimal</td>
<td>no</td>
<td>gamma</td>
<td>exch.</td>
<td>0.52</td>
<td>0.065</td>
<td>premature, lowbirthwt, extracardiac, univentricular, archobstruction, nonwhite</td>
</tr>
<tr>
<td>yes</td>
<td>none</td>
<td>minimal</td>
<td>yes</td>
<td>gamma</td>
<td>exch.</td>
<td>-0.08</td>
<td>0.51</td>
<td>premature, extracardiac, univentricular, archobstruction</td>
</tr>
<tr>
<td>yes</td>
<td>2 stage</td>
<td>minimal</td>
<td>yes</td>
<td>gamma</td>
<td>exch.</td>
<td>-1.83</td>
<td>0.02</td>
<td>premature, extracardiac, univentricular</td>
</tr>
<tr>
<td>yes</td>
<td>3 stage</td>
<td>minimal</td>
<td>yes</td>
<td>gamma</td>
<td>exch.</td>
<td>-0.34$^a$</td>
<td>0.22</td>
<td>lowbirthwt, extracardiac</td>
</tr>
<tr>
<td>yes</td>
<td>none</td>
<td>minimal</td>
<td>no</td>
<td>gamma</td>
<td>unstruct.</td>
<td>-0.16$^a$</td>
<td>0.86</td>
<td>lowbirthwt, extracardiac, univentricular</td>
</tr>
<tr>
<td>yes</td>
<td>3 stage</td>
<td>minimal</td>
<td>no</td>
<td>gamma</td>
<td>unstruct.</td>
<td>-0.16$^a$</td>
<td>0.86</td>
<td>lowbirthwt, extracardiac, univentricular</td>
</tr>
<tr>
<td>yes</td>
<td>none</td>
<td>minimal</td>
<td>yes</td>
<td>gamma</td>
<td>unstruct.</td>
<td>--$^b$</td>
<td>--</td>
<td>N/A</td>
</tr>
<tr>
<td>yes</td>
<td>2 stage</td>
<td>minimal</td>
<td>yes</td>
<td>gamma</td>
<td>unstruct.</td>
<td>0.14$^a$</td>
<td>0.3</td>
<td>extracardiac, univentricular, archobstruction</td>
</tr>
<tr>
<td>yes</td>
<td>3 stage</td>
<td>minimal</td>
<td>yes</td>
<td>gamma</td>
<td>unstruct.</td>
<td>0.68$^a$</td>
<td>0.58</td>
<td>lowbirthwt, extracardiac, univentricular, archobstruction</td>
</tr>
<tr>
<td>yes</td>
<td>none</td>
<td>minimal</td>
<td>no</td>
<td>Poisson</td>
<td>exch.</td>
<td>-0.08</td>
<td>0.23</td>
<td>premature, lowbirthwt, extracardiac, univentricular, archobstruction, nonwhite</td>
</tr>
<tr>
<td>yes</td>
<td>2 stage</td>
<td>minimal</td>
<td>no</td>
<td>Poisson</td>
<td>exch.</td>
<td>-0.46</td>
<td>0.4</td>
<td>premature, lowbirthwt, extracardiac, univentricular, archobstruction, nonwhite</td>
</tr>
<tr>
<td>yes</td>
<td>3 stage</td>
<td>minimal</td>
<td>no</td>
<td>Poisson</td>
<td>exch.</td>
<td>-0.49</td>
<td>0.31</td>
<td>premature, lowbirthwt, extracardiac, univentricular, archobstruction, nonwhite</td>
</tr>
<tr>
<td>yes</td>
<td>none</td>
<td>minimal</td>
<td>yes</td>
<td>Poisson</td>
<td>exch.</td>
<td>-0.08</td>
<td>0.21</td>
<td>premature, extracardiac, univentricular, archobstruction</td>
</tr>
<tr>
<td>yes</td>
<td>2 stage</td>
<td>minimal</td>
<td>yes</td>
<td>Poisson</td>
<td>exch.</td>
<td>-0.83</td>
<td>0.22</td>
<td>premature, extracardiac, univentricular, archobstruction</td>
</tr>
<tr>
<td>yes</td>
<td>3 stage</td>
<td>minimal</td>
<td>yes</td>
<td>Poisson</td>
<td>exch.</td>
<td>-0.63</td>
<td>0.17</td>
<td>premature, extracardiac, univentricular, archobstruction</td>
</tr>
</tbody>
</table>

All models are GLM models, using the SAS `genmod` procedure.

$^a$Met convergence criteria and provided parameter estimates but terminated in error; $^b$Terminated in error before providing parameter estimates.
APPENDIX B
CODE USED IN ANALYSES

SAS Code

/**********************************************************/
****** POTENTIAL COST BENEFIT FROM EARLY DETECTION OF ******
****** OF CRITICAL CONGENITAL HEART DISEASE **********/
/**********************************************************/

/*****************************/
******MACRO DEFINITIONS*****/
/*****************************/

data _null_;  
%LET categ_covars_all =race;  
%LET categ_stmt_all = class &categ_covars_all;  
%LET covars_all = female premature lowbirthwt multigest cesarean extracardiac univentricular archobstruction income poverty &categ_covars_all;  
%LET categ_covars_noincome =&categ_covars_all;  
%LET categ_stmt_noincome = class &categ_covars_noincome;  
%LET covars_noincome = female premature lowbirthwt multigest cesarean extracardiac univentricular archobstruction &categ_covars_noincome;  
%LET categ_covars_minimal =;  
%LET categ_stmt_minimal =;  
%LET covars_minimal = nonwhite premature lowbirthwt extracardiac univentricular archobstruction &categ_covars_minimal;  
%LET class_st_covar_prelim=&categ_stmt_noincome;  
%LET covar_list_prelim=&covars_noincome;  
%LET offset_var = logenrollmo;  
%LET offset_opt = offset=&offset_var;  
%put _all_;  
run;

/*****************************/
******DATA PREPARATION******/
/*****************************/

libname archer 'p:\data';  
%macro CLEAROUTPUTWINDOW;  
ods html close; /* close previous */  
ods html; /* open new */
%mend CLEAROUTPUTWINDOW;

%CLEAROUTPUTWINDOW;

options mprint;

/* Code the independent variables for analysis */
data transformeddata;
  set Archer.CHD_Final_v_2_23_13;

  /* lesion type */
  univentricular=0;
  archobstruction=0;
  if f_CHD_4=1 then univentricular=1; /*'single ventricle'*/
  if f_CHD_7=1 then univentricular=1; /*tricuspid atresia or stenosis*/
  if f_CHD_9=1 then archobstruction=1; /*aortic stenosis*/
  if f_CHD_10=1 then archobstruction=1; /*mitral stenosis*/
  if f_CHD_11=1 then archobstruction=1; /*HLHS*/
  if f_CHD_12=1 then archobstruction=1; /*subaortic stenosis*/
  if f_CHD_14=1 then archobstruction=1; /*Shone's complex*/
  if f_CHD_15=1 then archobstruction=1; /*coarc or IAA*/
  if f_CHD_16=1 then archobstruction=1; /*aortic atresia*/
  if f_CHD_16=1 then univentricular=1; /*aortic atresia*/

  /* binary variables with new names for coding clarity */
  earlydetect=0; earlydetect=1; earlydetect=.;
  if sex='M' then female=0; else if sex='F' then female=1; else female=.;

  /*additional transformations*/
  if exp_med <= 0 then logcost=.;
  else logcost=log(exp_med); /*log transform cost */
  enrollmo=months; /* prepare offset variable for modeling log cost*/
  logenrollmo = log(enrollmo);
  logenrollmocorr = log(enrollmocorr);
  difftime = min_cardsrg - min_hosptl; /* prepare the instrumental variable */

  if f_Prem_0=1 then premature=0; else premature=1;
  if premature=0 then premcat="37 or more wks";
  else if f_Prem_9=1 then premcat="35-36 wks";
  else if f_Prem_1=1 then premcat="Premature Unspecified";
  else if f_Prem_2=1 or f_Prem_3=1 then premcat="24 or less wks";
  else if f_Prem_4=1 then premcat="25-26 wks";
  else if f_Prem_5=1 then premcat="27-28 wks";
  else if f_Prem_6=1 then premcat="29-30 wks";
  else if f_Prem_7=1 then premcat="31-32 wks";
  else premcat="";

  if f_LBW_0=1 then lowbirthwt=0; else lowbirthwt=1;
if lowbirthwt=0 then birthwtcat = "E. NORMAL (>=2500g)";
else if f_LBW_7=1 or f_LBW_8=1 or f_LBW_9=1 /*or f_LBW_1=1*/ then
birthwtcat="D. LBW (1500-2499g)";
else if f_LBW_1=1 then birthwtcat="LBW Unspecified";
else if f_LBW_5=1 or f_LBW_6=1 then birthwtcat="C. VLBW (1000-1.499g)";
else if f_LBW_3=1 or f_LBW_4=1 then birthwtcat="B. ELBW (500-999g)";
else if f_LBW_2=1 then birthwtcat = "A. <500g";
else birthwtcat="";

format birthhospwithother $char50.;
if BLNGProv_Name='' then birthhospwithother='OTHER';
else birthhospwithother=BLNG_Prov_Name;

if race_txt='White' then nonwhite=0; else nonwhite=1;
format race $char25.;
if race_txt = '' then race='Unknown / Other';
else race=race_txt;

monthsoldonlastday = (age_last_day / 365) * 12;

rename BLNG_Prov_Name=birthhosp f_MG=multgest f_EA=extracardiac
f_CS=cesarean exp_med=rawcost crgsum_txt=crg crgsum=clinicalresourcegroup;
run;

/* Lesion classification*/
data transformeddata;
set transformeddata;
lesiontype=1;
if univentricular=0 and archobstruction=1 then lesiontype=2;
else if univentricular=1 and archobstruction=0 then lesiontype=3;
else if univentricular=1 and archobstruction=1 then lesiontype=4;
if f_CHD_15=1 then coarctation=1; else coarctation=0;
run;

/* Keep variables of interest and label them for clarity of output */
data cleandata;
set transformeddata;
keep rawcost logcost earlydetect female cesarean extracardiac
lowbirthwt premature multgest race nonwhite birthhosp
birthhospwithother
income poverty difftime enrollmo enrollmocorr univentricular
archobstruction lesiontype
logenrollmo logenrollmocorr difftime clinicalresourcegroup premcat
birthwtcat coarctation
monthsoldonlastday;
label rawcost='Raw Cost'
logcost='Natural Log of Raw Cost'
earlydetect='Early Detection'
female='Female'
cesarean='Cesarean Section'
extracardiac='Extracardiac Anomaly'
lowbirthwt='Low Birth Weight'
premature='Premature'
multgest='Multiple Gestation'
nonwhite='Non-white Race/Ethnicity'
birthhosp='Birth Hospital'
birthhospitalwithother='Birth Hospital (missing values grouped together)'
income='Median Income for Census Tract'
poverty='Percent of Children in Poverty for Census Tract'
difftime='Interhospital Driving Time (min)'
enrollmo='Months Enrolled (uncorrected)'
enrollmocorr='Months Enrolled (with administrative correction)'
logenrollmo='Ln(Months Enrolled-uncorrected)'
logenrollmocorr='Ln(Months Enrolled-with administrative correction)'
race='Race/Ethnicity'
univentricular='Single Ventricle'
archobstruction='Arch Obstruction'
coarctation='Aortic Arch Coarctation, Hyoplasia, or Interruption'
lesiontype='Lesion Type'
clinicalresourcegroup='Clinical Resource Group'
premcat='Prematurity Category'
birthwtcat='Birth Weight Category'
monthsoldonlastday='Age (mo) at End of Study';
run;

/* Export to STAT for selection bias procedure */
proc export data=cleandata outfile="P:\data\CHD_data";
run;

/***************************************************************************/
/*****BASIC DATA DIAGNOSTICS, NORMALITY TESTING, EXPLORATORY STATISTICS******/
/***************************************************************************/

/* Test normality for raw cost and ln(cost) */
/* -note: ignores months enrolled*/
title 'Expenditure Normality Testing - Overall';
ods graphics on;
ods select BasicMeasures ExtremeObs Quantiles Histogram Moments TestsForNormality ProbPlot;
proc univariate data=cleandata normaltest;
  var rawcost logcost;
  histogram rawcost / kernel(color=red) name='Overall Raw Cost Distribution';
    inset mean std / format=6.4;
  probplot rawcost / normal (mu=est sigma=est) square;
  histogram logcost / kernel(color=red) name='Overall Log Cost Distribution';
    inset mean std / format=6.4;
  probplot logcost / normal (mu=est sigma=est) square;
run;

title 'Expenditure Normality Testing by Group';
proc univariate data=cleandata normaltest;
  var rawcost logcost;
  class earlydetect;
  histogram rawcost / kernel(color=red) name='Raw Cost Distribution by Group';
inset mean std / format=6.4;
   probplot rawcost / normal (mu=est sigma=est) square;
histogram logcost / kernel(color=red) name='Log Cost Distribution by Group';
inset mean std / format=6.4;
   probplot logcost / normal (mu=est sigma=est) square;
run;

/* Diagnostic data for numerical covariates */
title 'Covariate Baseline Diagnostics';
proc univariate data=cleandata normaltest;
   var difftime income poverty enrollmo clinicalresourcegroup;
   histogram difftime / kernel(color=red);
inset mean std / format=6.4;
   probplot difftime / normal (mu=est sigma=est) square;
   histogram income / kernel(color=red);
inset mean std / format=6.4;
   probplot income / normal (mu=est sigma=est) square;
   histogram poverty / kernel(color=red);
inset mean std / format=6.4;
   probplot poverty / normal (mu=est sigma=est) square;
   histogram enrollmo / kernel(color=red);
inset mean std / format=6.4;
   probplot enrollmo / normal (mu=est sigma=est) square;
   histogram clinicalresourcegroup / kernel(color=red);
inset mean std / format=6.4;
run;

proc ttest data=cleandata cochran ci=equal umpu;
   class earlydetect;
   var rawcost logcost difftime income poverty enrollmo clinicalresourcegroup;
run;
quit;

proc freq data=cleandata order=formatted;
   tables female / plots(only)=freqplot(scale=percent);
   tables earlydetect*female / chisq cmh plots(only)=freqplot(scale=percent);
   tables race / plots(only)=freqplot(scale=percent);
   tables earlydetect*race / chisq cmh plots(only)=freqplot(scale=percent);
   tables nonwhite / plots(only)=freqplot(scale=percent);
   tables earlydetect*nonwhite / plots(only)=freqplot(scale=percent);
   tables premature / plots(only)=freqplot(scale=percent);
   tables earlydetect*premature / chisq cmh plots(only)=freqplot(scale=percent);
   tables premcat / plots(only)=freqplot(scale=percent);
   tables earlydetect*premcat / chisq cmh plots(only)=freqplot(scale=percent);
   tables lowbirthwt / plots(only)=freqplot(scale=percent);
   tables earlydetect*lowbirthwt / chisq cmh plots(only)=freqplot(scale=percent);
   tables birthwtcat / plots(only)=freqplot(scale=percent);
   tables earlydetect*birthwtcat / chisq cmh plots(only)=freqplot(scale=percent);
   tables multgest / plots(only)=freqplot(scale=percent);
   tables earlydetect*multgest / chisq cmh plots(only)=freqplot(scale=percent);
tables cesarean / plots(only)=freqplot(scale=percent);
tables earlydetect*cesarean / chisq cmh
plots(only)=freqplot(scale=percent);
tables archobstruction / plots(only)=freqplot(scale=percent);
tables earlydetect*archobstruction / chisq cmh
plots(only)=freqplot(scale=percent);
tables coarctation / plots(only)=freqplot(scale=percent);
tables earlydetect*coarctation / chisq cmh
plots(only)=freqplot(scale=percent);
tables univentricular / plots(only)=freqplot(scale=percent);
tables earlydetect*univentricular / chisq cmh
plots(only)=freqplot(scale=percent);
tables clinicalresourcegroup / plots(only)=freqplot(scale=percent);
tables earlydetect*clinicalresourcegroup / chisq cmh
plots(only)=freqplot(scale=percent);
tables extracardiac / plots(only)=freqplot(scale=percent);
tables earlydetect*extracardiac / chisq cmh
plots(only)=freqplot(scale=percent);
run;
quit;

title 'Variable Differences by Race/Ethnicity';
proc anova data=cleandata;
class race;
model rawcost logcost difftime income poverty enrollmo
clinicalresourcegroup = race;
run;
quit;

proc freq data=cleandata order=formatted;
tables race*female / chisq cmh plots(only)=freqplot(scale=percent);
tables race*premature / chisq cmh plots(only)=freqplot(scale=percent);
tables race*premcat / chisq cmh plots(only)=freqplot(scale=percent);
tables race*lowbirthwt / chisq cmh plots(only)=freqplot(scale=percent);
tables race*birthwtcat / chisq cmh plots(only)=freqplot(scale=percent);
tables race*multigest / chisq cmh plots(only)=freqplot(scale=percent);
tables race*cesarean / chisq cmh plots(only)=freqplot(scale=percent);
tables race*archobstruction / chisq cmh
plots(only)=freqplot(scale=percent);
tables race*coarctation / chisq cmh plots(only)=freqplot(scale=percent);
tables race*univentricular / chisq cmh
plots(only)=freqplot(scale=percent);
tables race*clinicalresourcegroup / chisq cmh
plots(only)=freqplot(scale=percent);
run;
quit;

title 'Variable differences by White/Non-white Race/Ethnicity';
proc ttest data=cleandata cochran ci=equal umpu;
class nonwhite;
var rawcost logcost difftime income poverty enrollmo
clinicalresourcegroup;
run;
quit;

proc freq data=cleandata order=formatted;
tables nonwhite*female / chisq cmh plots(only)=freqplot(scale=percent);
tables nonwhite*nonwhite / plots(only)=freqplot(scale=percent);
tables nonwhite*premature / chisq cmh plots(only)=freqplot(scale=percent);
tables nonwhite*premcat / chisq cmh plots(only)=freqplot(scale=percent);
tables nonwhite*lowbirthwt / chisq cmh plots(only)=freqplot(scale=percent);
tables nonwhite*birthwtcat / chisq cmh plots(only)=freqplot(scale=percent);
tables nonwhite*multgest / chisq cmh plots(only)=freqplot(scale=percent);
tables nonwhite*cesarean / chisq cmh plots(only)=freqplot(scale=percent);
tables nonwhite*archobstruction / chisq cmh plots(only)=freqplot(scale=percent);
tables nonwhite*coarctation / chisq cmh plots(only)=freqplot(scale=percent);
tables nonwhite*univentricular / chisq cmh plots(only)=freqplot(scale=percent);
tables nonwhite*clinicalresourcegroup / chisq cmh plots(only)=freqplot(scale=percent);
tables nonwhite*multgest / chisq cmh plots(only)=freqplot(scale=percent);
tables nonwhite*cesarean / chisq cmh plots(only)=freqplot(scale=percent);
tables nonwhite*archobstruction / chisq cmh plots(only)=freqplot(scale=percent);
tables nonwhite*coarctation / chisq cmh plots(only)=freqplot(scale=percent);
tables nonwhite*univentricular / chisq cmh plots(only)=freqplot(scale=percent);
tables nonwhite*clinicalresourcegroup / chisq cmh plots(only)=freqplot(scale=percent);
tables nonwhite*multgest / chisq cmh plots(only)=freqplot(scale=percent);
tables nonwhite*cesarean / chisq cmh plots(only)=freqplot(scale=percent);
tables nonwhite*archobstruction / chisq cmh plots(only)=freqplot(scale=percent);
tables nonwhite*coarctation / chisq cmh plots(only)=freqplot(scale=percent);
tables nonwhite*univentricular / chisq cmh plots(only)=freqplot(scale=percent);
tables nonwhite*clinicalresourcegroup / chisq cmh plots(only)=freqplot(scale=percent);
run;
quit;

title 'Frequency listing of birth hospitals';
proc freq data=cleandata order=freq;
  tables birthhosp birthhospwithother;
run;
quit;

title 'Relationship Between Months Enrolled and Age at End of Study';
proc reg data=cleandata noprint;
  model enrollmo = monthsoldonlastday;
  plot enrollmo*monthsoldonlastday;
run;
quit;

/**************************************************
 ******/ TEST INTRACLASS CORRELATION COEFFICIENT ******/
/***************************************************/

/* Intraclass correlation coefficient code without grouping missing values */
title 'ICC (Missing Values Not Grouped, Method 1)';
ods output CovParms = covp;
proc mixed data = cleandata;
  class birthhosp;
  model rawcost = ;
  random intercept /subject= birthhosp;
run;

data icc;
  set covp end=last;
  retain bvar;
  if subject='' then bvar = estimate;
  if last then icc = bvar/(bvar+estimate);
run;
proc print data = icc;
run;
title 'ICC (Missing Values Not Grouped, Method 2)';
proc mixed ratio data = cleandata;
   class birthhosp;
   model rawcost = ;
   random intercept /subject= birthhosp;
run;

/* Intraclass correlation coefficient code WITH grouping missing values */
title 'ICC (Missing Values Grouped, Method 1)';
ods output CovParms = covp;
proc mixed data = cleandata;
   class birthhospwithother;
   model rawcost = ;
   random intercept /subject= birthhospwithother;
run;
data icc;
   set covp end=last;
   retain bvar;
   if subject~="" then bvar = estimate;
   if last then icc = bvar/(bvar+estimate);
run;
proc print data = icc;
run;

/* The next sections prepare for Park test for heteroskedasticity using two methods */

/**************************
/*****TEST HETEROSKEDASTICITY*****
/**************************/

title 'ICC (Missing Values Grouped, Method 2)';
proc mixed ratio data = cleandata;
   class birthhospwithother;
   model rawcost = ;
   random intercept /subject= birthhospwithother;
run;
quit;

/**************************
/*****TEST HETEROSKEDASTICITY*****
/**************************/

*OLS model using proc glm to generate residuals and predicted values on log transformed cost *
/* note: need offset equivalent here */
title 'OLS Model to Generate Residuals';
proc glm data=cleandata;
   &class_st_covar_prelim;
   model logcost = earlydetect &covar_list_prelim &offset_var / solution;
   output out=residualdatabyols
      p=logcostthat
      r=logcostresid
     stdr=stderrofresid;
run;
quit;
/* Generalized linear models using proc genmod to generate predicted values of cost using log link */

/* GLM with Poisson distribution */
/* note: the predicted value is re-exponentiated */
/* note: need to specify covariance matrix */
title 'GLM (Poisson distribution) to Generate Residuals';
proc genmod data=cleandata;
  &class_st_covar_prelim;
  model rawcost = earlydetect &covar_list_prelim
    / link=log
dist=poisson
    &offset_opt;
  output out=residualdatabyglmpoisson
    p=costhatpoisson;
run;
quit;

/* Create Park test variables (OLS method) */
data residualdatabyols;
  set residualdatabyols;
  yhati = exp(logcosthat + (0.5 * (stderrofresid ** 2)));
parkpred = log(yhati);
parkpredid = log((rawcost - yhati) ** 2);
  label parkresid='Ln(Yi-Yhati)^2'
parkpred='Ln(Yhati)';
run;

/* Create Park test variables for GLM with Poisson distribution */
data residualdatabyglmpoisson;
  set residualdatabyglmpoisson;
parkpred = log(costhatpoisson);
parkresid = log((rawcost - costhatpoisson) ** 2);
  label parkresid='Ln(Yi-Yhati)^2'
parkpred='Ln(Yhati)';
run;

/* Create Park test variables for GLM with gamma distribution */
data residualdatabyglmgamma;
  set residualdatabyglmgamma;
parkpred = log(costhatgamma);
run;
parkresid = log((rawcost - costhatgamma) ** 2);
label parkresid='Ln(Yi-Yhati)^2'
parkpred='Ln(Yhati)';
run;

/* Park test by OLS */
title 'Park Test for Heteroskedasticity (OLS model)';
proc reg data=residualdatabyols noprint;
  model parkresid = parkpred;
  plot parkresid*parkpred;
run;
quit;

/* Park test by GLM with Poisson distribution */
title 'Park Test for Heteroskedasticity (GLM with Poisson distribution)';
proc reg data=residualdatabyglmpoisson noprint;
  model parkresid = parkpred;
  plot parkresid*parkpred;
run;
quit;

/* Park test by GLM with gamma distribution */
title 'Park Test for Heteroskedasticity (GLM with Gamma distribution)';
proc reg data=residualdatabyglmgamma noprint;
  model parkresid = parkpred;
  plot parkresid*parkpred;
run;
quit;

/****************************/
/*****PREPARE INSTRUMENTAL VARIABLE*****/
/****************************/

/* Instrumental variable model - Logistic Regression with following parameters:
 inputdataset
 predictorvar: left hand side of equation
 categ_class_stmt: name of macro for class statement for categorical variables; should be null if none
 covariate_list: covariate list including categoricals;
 outputdataset;
 outputvariable: name of variable predicted probability
*/
%macro IVMODEL(inputdataset,predictorvar,categ_class_stmt,covariate_list,outputdataset,outputvariable);
  proc logistic data=&inputdataset;
    &categ_class_stmt;
    model earlydetect (ref=first) = &predictorvar &covariate_list / link=logit;
    output out=&outputdataset p=&outputvariable;
  run;
  quit;
%end;
%mend IVMODEL;

title 'First stage Instrumental Variable preparation - Maximum Model';
%IVMODEL(cleandata,difftime,&categ_stmt_all,&covars_all,iv1_max,predictorvar_phi);

title 'Second Stage Instrumental Variable preparation - Maximum Model';
%IVMODEL(iv1_max,predictorvar_phi,&categ_stmt_all,&covars_all,iv2_max,predictorvar_w);

title 'First stage Instrumental Variable preparation - Maximum Model without Income/Poverty';
%IVMODEL(cleandata,difftime,&categ_stmt_noincome,&covars_noincome,iv1_noinc,predictorvar_phi);

title 'Second Stage Instrumental Variable preparation - Maximum Model without Income/Poverty';
%IVMODEL(iv1_noinc,predictorvar_phi,&categ_stmt_noincome,&covars_noincome,iv2_noinc,predictorvar_w);

title 'First stage Instrumental Variable preparation - Minimum Model';
%IVMODEL(cleandata,difftime,&categ_stmt_minimal,&covars_minimal,iv1_min,predictorvar_phi);

title 'Second Stage Instrumental Variable preparation - Minimum Model';
%IVMODEL(iv1_min,predictorvar_phi,&categ_stmt_minimal,&covars_minimal,iv2_min,predictorvar_w);

title 'First stage Instrumental Variable preparation - No Covariates';
%IVMODEL(cleandata,difftime,,,iv1_min,predictorvar_phi);

title 'Second Stage Instrumental Variable preparation - No Covariates';
%IVMODEL(iv1_min,predictorvar_phi,,,iv2_min,predictorvar_w);

/****************************/
/*****FINAL MODELING*****/
/****************************/

/* Generalized Linear Model (no random effects) /GEE with Parameters: */
inputdata
  multilevel_flag: 1 if accounting for cluster, 0 if not
  clustervar: name of clustering var, insert dummy if not
  selectionvar: the selection variable or IV-predicted probability
  categ_class_stmt: name of macro for class statement for categorical
  variables; should be null if none
  covariate_list: covariate list including categoricals;
  offset_flag: 1 if there is an offset variable; 2 if it goes in the
  covariate list; 0 if none;
  offset_name: in form offset=offsetvariable;
  distributiontype: typically gamma or poisson
  covarstruct: IND, CS (exchangeable), UN, AR(1);
*/

%macro GENLINEARMODEL(inputdata,multilevel_flag,clustervar,selectionvar,categ_class_stmt,covariate_list,offset_flag,offset_name,distributiontype,covarstruct);
proc genmod data=&inputdata;
  &categ_class_stmt;
  %if &multilevel_flag=1 %then class &clustervar;
%end;

91
model rawcost = &selectionvar &covariate_list %if &offset_flag=2 %then &offset_name;
   / link=log
   dist=&distributiontype
   %if &offset_flag=1 %then offset=&offset_name;
   type3 maxiter=100 corrb covb;
   %if &multilevel_flag=1 %then repeated subject = &clustervar / corr=&covarstruct;
run;
quit;
%mend GENLINEARMODEL;

title 'Generalized Linear Model with Multilevel Modeling - Maximal Model - No IV, offset as covariate, gamma distribution, exchangeable covariance matrix';
%GENLINEARMODEL(cleandata,1,birthhospwithother,earlydetect,&categ_stmt_all,&covars_all,2,&offset_var,gamma,CS);
title 'Generalized Linear Model with Multilevel Modeling - Maximal Model - Two-Stage IV, offset as covariate, gamma distribution, exchangeable covariance matrix';
%GENLINEARMODEL(iv1_noinc,1,birthhospwithother,predictorvar_phi,&categ_stmt_all,&covars_all,2,&offset_var,gamma,CS);

%GENLINEARMODEL(iv2_noinc,1,birthhospwithother,predictorvar_w,&categ_stmt_all,&covars_all,2,&offset_var,gamma,CS);

title 'Generalized Linear Model WITHOUT Multilevel Modeling - Minimal Model - No IV, offset as covariate, gamma distribution, exchangeable covariance matrix';
%GENLINEARMODEL(cleandata,0,birthhospwithother,earlydetect,&categ_stmt_minimal,&covars_minimal,2,&offset_var,gamma,CS);

title 'Generalized Linear Model WITHOUT Multilevel Modeling - Minimal Model - Two-Stage IV, offset as covariate, gamma distribution, exchangeable covariance matrix';
%GENLINEARMODEL(iv1_noinc,0,birthhospwithother,predictorvar_phi,&categ_stmt_minimal,&covars_minimal,2,&offset_var,gamma,CS);

title 'Generalized Linear Model WITHOUT Multilevel Modeling - Minimal Model - Three-Stage IV, offset as covariate, gamma distribution, exchangeable covariance matrix';
%GENLINEARMODEL(iv2_noinc,0,birthhospwithother,predictorvar_w,&categ_stmt_minimal,&covars_minimal,2,&offset_var,gamma,CS);

%GENLINEARMODEL(cleandata,1,birthhospwithother,earlydetect,&categ_stmt_minimal,&covars_minimal,2,&offset_var,gamma,CS);

%GENLINEARMODEL(iv1_noinc,1,birthhospwithother,predictorvar_phi,&categ_stmt_minimal,&covars_minimal,2,&offset_var,gamma,CS);

%GENLINEARMODEL(iv2_noinc,1,birthhospwithother,predictorvar_w,&categ_stmt_minimal,&covars_minimal,2,&offset_var,gamma,CS);

%GENLINEARMODEL(cleandata,1,birthhospwithother,earlydetect,&categ_stmt_minimal,&covars_minimal,2,&offset_var,gamma,CS);

%GENLINEARMODEL(iv1_noinc,1,birthhospwithother,predictorvar_phi,&categ_stmt_minimal,&covars_minimal,2,&offset_var,gamma,CS);

%GENLINEARMODEL(iv2_noinc,1,birthhospwithother,predictorvar_w,&categ_stmt_minimal,&covars_minimal,2,&offset_var,gamma,CS);
%GENLINEARMODEL(iv2_noinc,1,birthhospwithother,predictorvar_w,&categ_stmt_minimal,&covars_minimal,2,&offset_var,gamma,CS);

title 'Generalized Linear Model with Multilevel Modeling - Minimal Model - No IV, offset as offset, gamma distribution, exchangeable covariance matrix';
%GENLINEARMODEL(cleandata,1,birthhospwithother,earlydetect,&categ_stmt_minimal,1,&offset_var,gamma,CS);

%GENLINEARMODEL(iv1_noinc,1,birthhospwithother,predictorvar_phi,&categ_stmt_minimal,&covars_minimal,1,&offset_var,gamma,UN);

%GENLINEARMODEL(iv2_noinc,1,birthhospwithother,predictorvar_w,&categ_stmt_minimal,&covars_minimal,1,&offset_var,gamma,UN);

%GENLINEARMODEL(cleandata,1,birthhospwithother,earlydetect,&categ_stmt_minimal,1,&offset_var,poisson,CS);

%GENLINEARMODEL(iv1_noinc,1,birthhospwithother,predictorvar_phi,&categ_stmt_minimal,1,&offset_var,poisson,UN);

%GENLINEARMODEL(iv2_noinc,1,birthhospwithother,predictorvar_w,&categ_stmt_minimal,1,&offset_var,poisson,UN);

%GENLINEARMODEL(cleandata,1,birthhospwithother,earlydetect,&categ_stmt_minimal,1,&offset_var,poisson,CS);
%GENLINEARMODEL(iv1_noinc,1,birthhospwithother,predictorvar_phi,&categ_stmt_minimal,&covars_minimal,2,&offset_var,poisson,CS);
title 'Generalized Linear Model with Multilevel Modeling - Minimal Model - Three-Stage IV, offset as covariate, poisson distribution, exchangeable covariance matrix';
%GENLINEARMODEL(iv2_noinc,1,birthhospwithother,predictorvar_w,&categ_stmt_minimal,&covars_minimal,2,&offset_var,poisson,CS);
title 'Generalized Linear Model with Multilevel Modeling - Minimal Model - No IV, offset as offset, poisson distribution, exchangeable covariance matrix';
%GENLINEARMODEL(cleandata,1,birthhospwithother,earlydetect,&categ_stmt_minimal,&covars_minimal,1,&offset_var,poisson,CS);
title 'Generalized Linear Model with Multilevel Modeling - Minimal Model - Two-Stage IV, offset as offset, poisson distribution, exchangeable covariance matrix';
%GENLINEARMODEL(iv1_noinc,1,birthhospwithother,predictorvar_phi,&categ_stmt_minimal,&covars_minimal,1,&offset_var,poisson,CS);
title 'Generalized Linear Model with Multilevel Modeling - Minimal Model - Three-Stage IV, offset as offset, poisson distribution, exchangeable covariance matrix';
%GENLINEARMODEL(iv2_noinc,1,birthhospwithother,predictorvar_w,&categ_stmt_minimal,&covars_minimal,1,&offset_var,poisson,CS);

/*************/
/*****END*****
/*******END*****/
STATA Code

use "P:\Data\CHD_data.dta", clear

treatreg logcost nonwhite premature lowbirthwt extracardiac univentricular archobstruction logenrollmo, treat(earlydetect = difftime female multgest cesarean)
.

REFERENCES


51. Peiris V, Singh TP, Tworetzky W, Chong EC, Gauvreau K, Brown DW. Association of socioeconomic position and medical insurance with fetal diagnosis of


54. Sebelius KT. 2011;


60. Pulse oximetry screening in newborns: a policy position from the American Heart Association: American Heart Association; June 2012.


BIOGRAPHICAL SKETCH

The author was born in Delaware and grew up in Gainesville, FL. He received his Bachelor of Arts, with highest honors, from Princeton University in 1998, with an additional Certificate in Applications of Computing. He earned his Doctor of Medicine, with honors, from the University of Florida College of Medicine in 2004. He completed residency training in Pediatrics at the University of Vermont in 2009, and served there as Chief Resident from 2009-2010. He is currently completing a fellowship in Pediatric Cardiology and a Master of Science degree with a concentration in Health Outcomes and Policy at the University of Florida. He lives in Gainesville, Florida with his wife and three wonderful children. His future plans involve Pediatric Cardiology practice in Montana with continued involvement in child health policy and research.