COST-EFFECTIVENESS AND VALUE OF INFORMATION ANALYSES OF SCREEN-AND-TREAT STRATEGIES FOR OPIOID USE DISORDER IN PREGNANCY

A Dissertation
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA BY

Viengneesee Thao

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Karen M. Kuntz, ScD; Eva A. Enns, PhD

August 2020
Acknowledgements

There are many people I’d like to thank for supporting me throughout my graduate studies.

I first want to thank my husband, Donny Xiong. Donny has read every iteration of my dissertation without complaint. His editorial services are greatly appreciated. Next, I’d like to thank my mom and brothers: Doua, Yeng Kong, Leimeang, Zenki, and Leepeng, for their love, support, and encouragement.

I also thank my peers, Dr. Szu-Yu (Zoe) Kao and Dr. Fernando Alarid-Escudero who have answered every question I have had about modeling and programming and have never hesitated to help. I am a better researcher because I was able to observe and learn from them.

My academic advisors, Dr. Karen M. Kuntz and Dr. Eva A. Enns and committee members, Dr. Katy B. Kozhimannil and Dr. Robert Levy have been patient and given me valuable feedback and advice. I thank them and think it an honor to have been given the opportunity to work with them.

This dissertation was supported by the Doctoral Dissertation Fellowship which was possible because of the recommendation letters Dr. Karen M. Kuntz and Dr. John A. Nyman wrote; thank you, both. I also thank Dr. Bryan Dowd for securing funding from the Agency for Healthcare Research & Quality T32 training grant program to support my studies.
# Table of Contents

Acknowledgements.................................................................................................................. i
Table of Contents...................................................................................................................... ii
List of Tables ............................................................................................................................... v
List of Figures ............................................................................................................................. vi
List of Abbreviations ................................................................................................................ vii

Chapter 1 Introduction................................................................................................................. 8

Chapter 2 Cost-Effectiveness Strategy to Screen-and-Treat Pregnant Women with Opioid Use Disorder .............................................................................................................. 12

2.1 Introduction .......................................................................................................................... 12

2.2 Methods .............................................................................................................................. 13
  2.2.1 Screening for OUD ........................................................................................................ 13
  2.2.2 Treatment for OUD ....................................................................................................... 14
  2.2.3 Markov Models ............................................................................................................ 16
  2.2.4 Transition Probabilities ............................................................................................... 18
  2.2.5 Health Sector Perspective ........................................................................................... 19
  2.2.6 Cost of Screening ....................................................................................................... 19
  2.2.7 Cost of Treatment ....................................................................................................... 20
  2.2.7 Health Services Utilization and Cost ......................................................................... 20
  2.2.8 Health-Related Quality of Life ................................................................................... 21
  2.2.9 Markov Simulation ...................................................................................................... 21
  2.2.9 Cost-Effectiveness Analysis ....................................................................................... 21
  2.2.9 Additional Analysis .................................................................................................... 22
  2.2.9 Sensitivity Analysis .................................................................................................... 22

2.3 Results ............................................................................................................................... 22
  2.3.1 Cost-Effectiveness Analysis ....................................................................................... 22
  2.3.2 Additional Analysis ................................................................................................... 24
  2.3.3 Sensitivity Analysis ................................................................................................... 24
2.4 Discussion .......................................................................................................................... 30
  2.4.1 Societal Perspective ................................................................................................. 31
  2.4.1 Limitations ............................................................................................................. 32
  2.4.2 Conclusion ............................................................................................................. 33

Chapter 3 Future Research on Treatment for Opioid Use Disorder in Pregnancy: A Value of Information Analysis .......................................................................................................................... 35
  3.1 Introduction .................................................................................................................. 35
  3.2 Methods ........................................................................................................................ 37
    3.2.1 Markov Model Review ......................................................................................... 38
    3.2.2 Probabilistic Sensitivity Analysis ......................................................................... 40
    3.2.3 Expected Value of Perfect Information ............................................................. 40
    3.2.4 Expected Value of Partial Perfect Information ................................................ 41
    3.2.5 Expected Value of Sample Information and Expected Net Benefit of Sampling ................................................................................................................................. 42
  3.3 Results ........................................................................................................................... 43
    3.3.1 Probabilistic Sensitivity Analysis ......................................................................... 43
    3.3.2 Expected Value of Population Perfect Information ........................................ 46
    3.3.3 Expected Value of Population Partial-Perfect Information ................................ 46
    3.3.4 Expected Value of Sample Information for Parameters (EVSI) and Expected Net Benefit of Sampling (ENBS) ................................................................................................................................. 49
  3.4 Discussion ..................................................................................................................... 52
    3.4.1 Limitations ........................................................................................................... 52
    3.4.2 Conclusion .......................................................................................................... 53

Chapter 4 Cost Estimators for Censored Data and Their Impact on Cost-Effectiveness Analyses ................................................................................................................................. 54
  4.1 Introduction ................................................................................................................... 54
  4.2 Methods ......................................................................................................................... 55
    4.2.1 Simulation ............................................................................................................ 55
List of Tables

Table 2.1 Parameters.................................................................................................................. 15
Table 2.2 Cost, Quality-Adjusted Life Year, and Incremental Cost-Effectiveness Ratio 23
Table 2.3 Adverse Outcomes Per 10,000 OUD-Related Pregnancies after 40 weeks, N( reduction) ................................................................................................................................. 25

Table 3.1 Parameters.................................................................................................................. 39

Table 4.1 Notation and Explanation ........................................................................................... 58
Table 4.2 Computing the Kaplan-Meier Estimate for Censoring ............................................... 61
Table 4.3 2-Year Cost-Effectiveness Results ............................................................................. 74
List of Figures

Figure 2.1 Maternal and Infant Markov Model .......................................................... 17
Figure 2.2 1-Way Sensitivity Analysis (4P twice + MMT vs 4P once + MMT).............. 26
Figure 2.3 2-Way Sensitivity Analysis ........................................................................ 28
Figure 2.4 Cost-Effectiveness Acceptability Curve .................................................... 29

Figure 3.1 Incremental Cost-Effectiveness Ratio of SURP + MMT v. 4P + MMT ........ 44
Figure 3.2 Cost-Effectiveness Acceptability Curve ..................................................... 45
Figure 3.3 Expected Value of Population Perfect Information ..................................... 47
Figure 3.4 Expected Value of Population Partial-Perfect Information on Screening with the 4Ps Plus, Substance Use Risk Profile, and Drug Abuse Screening Tool ................. 47
Figure 3.5 Expected Value of Population Partial-Perfect Information on Utility on Methadone and Buprenorphine Maintenance Treatment ........................................... 48
Figure 3.6 Expected Value of Population Partial-Perfect Information on Maternal and Neonatal Outcomes Associated With Timing of Relapse on and off Treatment .......... 48
Figure 3.7 Expected Net Benefit of Sampling for Screening with the 4Ps Plus, Substance Use Risk Profile, and Drug Abuse Screening Tool at WTP $100,000 ....................... 50
Figure 3.8 Expected Net Benefit of Utility on Methadone and Buprenorphine Maintenance Treatment at WTP $100,000 ................................................................. 50
Figure 3.9 Expected Net Benefit of Sampling for Maternal and Neonatal Outcomes Associated With Timing of Relapse on and off Treatment at WTP $100,000 .............. 51
Figure 3.10 Expected Value of Sample Information for Maternal and Neonatal Outcomes Associated with Timing of Relapse on and off Treatment ................................ 51

Figure 4.1 Cumulative 2-Year Cost Distribution of Treatment A and B ......................... 65
Figure 4.2 Cost Estimation Techniques’ Performance (Treatment A) .......................... 68
Figure 4.3 Cost Estimation Techniques’ Performance (Treatment B) .......................... 69
Figure 4.4 Survival Analysis ....................................................................................... 71
List of Abbreviations

4P 4Ps Plus screening tool
AS Average sample estimator
BMT Buprenorphine maintenance treatment
BT Bang and Tsiatis estimator
CC Complete case estimator
COR Cost of research
DAST Drug abuse screening tool
ENBS Expected net benefit of sampling
EVPI Expected value of perfect information
EVPPI Expected value of partial perfect information
EVSIE Expected value of sample information
HrQoL Health-related quality of life
ICER Incremental cost-effectiveness ratio
MMT Methadone maintenance treatment
MOTHER Maternal Opioid Treatment: Human Experimental Research Trial
NAS Neonatal abstinence syndrome
NMB Net monetary benefit
OUD Opioid use disorder
PSA Probabilistic sensitivity analysis
QALY Quality-adjusted life year
RCT Randomized control trial
SA Sensitivity analysis
SURP Substance use risk profile
VOI Value of information
WTP Willingness-to-pay
ZT Zhao and Tian estimator
Chapter 1 Introduction

Decision science is a collection of quantitative techniques, such as simulation modeling and cost-effectiveness analyses, used to inform decision making. In healthcare, decision science models may be used to regulate reimbursement in order to optimize value, to assess how cost-effective practice guidelines are, or to assess community-wide interventions. For example, in the United Kingdom, the National Health Service will not pay for a novel therapy until it is shown to be just as cost-effective as the standard of care.1 Another example of how a different country, the United States, uses decision science can be seen in recent publications that use cost-effectiveness analyses to assess new screening guidelines for specific types of breast and colorectal cancers. Currently, a simulation model is being used to assess whether Minnesota’s “stay-at-home” order was effective in reducing the spread of COVID-19.2-4

Decision models on best practices for treating substance abuse conditions are rare. This is especially true for opioid use disorder, particularly among pregnant women. There may be many reasons for this, including that pregnant women are often excluded from clinical trials due to safety concerns. Furthermore, women's health is often not a priority for policymakers, nor is the health of individuals who suffer from substance use disorders. Such exclusions result in difficulties in modeling how interventions could affect pregnant women because estimates are reflective only of the populations who are included in research or prioritized by policymakers. This means that few studies have explored the tradeoffs of available therapies for opioid use disorder in pregnancy.
Concern over increased opioid use disorder resulting in decreased life expectancy in the US has highlighted how problematic this illness is and brings new opportunities to inform health care policy and practice. In the next chapter of this dissertation, I conduct one of the first cost-effectiveness analyses on screening for and treating opioid use disorder in pregnancy. My model was informed by data from the Maternal Opioid Treatment: Human Experimental Research (MOTHER) trial, the first and only randomized control trial which assessed methadone and buprenorphine maintenance therapy in pregnancy. Additionally, I compare three screening tools along with treatment.

The next chapter is formatted for publication and is meant to be read as a standalone paper. Policymakers who wish to address the growing rates of maternal opioid use may find this chapter helpful. Providers looking for additional information on how various combinations of screening and pharmacotherapies affect pregnant women may also find this chapter helpful.

In the third chapter, I use the decision science technique known as a value of information analysis to determine if additional research regarding screening and treatment of opioid use disorder in pregnancy is valuable. New research may be warranted because model parameters are uncertain. Specifically, I opted to explore the following: 1) screening performance; 2) quality of life on methadone and buprenorphine; and 3) treatment effectiveness. To collect additional information on these parameters, I proposed two study designs, a randomized control trial and an observational study. In the value of information analysis, I compare the per-person cost of research under each study design to identify an optimal sample size for when additional research is warranted. These results may be helpful in setting research agendas, for securing grants for research,
or justifying why a decision maker would want to allocate funds to conduct additional research.

In the fourth chapter, I explore the use of cost estimators which correct for censored costs (i.e., incomplete follow-up data). Censored data is problematic for cost and cost-effectiveness evaluations because downstream costs or potential cost-savings are not observed. Follow-up costs for postpartum women taken from public health insurance datasets are subject to high amounts of censoring because public insurance programs often only cover prenatal, labor, and a portion of postpartum care. Given that pregnant women are especially prone to being censored from public health insurance data when they no longer qualify for coverage, challenges exist for evaluating screening strategies for opioid use disorder as all of the costs to screen are incurred at the beginning of pregnancy but some of the cost-savings are realized after birth and may be unobserved. To account for unobserved follow-up costs, I examine the impact of the Bang and Tsiatis and Zhao and Tian cost estimators, along with the average cost and complete case estimators in a simulated cost dataset to determine: 1) how censoring undervalues downstream cost estimates; 2) whether cost estimators are successful in producing less biased cost estimates; and 3) how adjusted cost estimates affect cost-effectiveness analyses. Health services researchers may find this information useful when working with censored cost data.

This dissertation utilizes decision science techniques to inform health care policy and practice regarding the increasing epidemic of opioid use disorder, specifically among pregnant women. Chapter two aims to identify a cost-effective screen-and-treat strategy for opioid use disorder; chapter three aims to identify areas of additional research that are
warranted; and chapter four explores cost estimators and how they affect cost and cost-effectiveness analyses. Taken together, these three chapters may be used as a practical guide on how to implement decision sciences to inform decision making.
Chapter 2 Cost-Effectiveness Strategy to Screen-and-Treat Pregnant Women with Opioid Use Disorder

2.1 Introduction

Opioid use disorder (OUD) is an illness with serious consequences that is increasingly common in the United States. As OUD continues to rise across the country, one specific population at risk is pregnant women. Between 2000 and 2009, OUD during pregnancy increased almost 4-fold. In 2017, 50,000 pregnant women reported they misused prescription opioids in the past 30 days. Pregnant women with OUD risk fetal loss, preterm birth, and maternal complications during delivery, including cardiac arrest and death. Furthermore, 68% of pregnant women with OUD will give birth to infants with neonatal abstinence syndrome (NAS), “a drug withdrawal syndrome that most commonly occurs in infants after in utero exposure to opioids” that costs the healthcare system $316 million annually.

For some women, prenatal care is the first and often only time when they receive routine clinical care. Therefore, screening for nonmedical opioid use during prenatal visits is an opportune time to identify and refer women for treatment. Universal screening for OUD among pregnant women is recommended by the American Society of Addiction Medicine. Screening tools that have been shown to be effective in identifying problematic substance use in pregnant women include the 4Ps Plus (4P) and the Substance-Use Risk Profile (SURP). The 4P and SURP have the potential to correctly identify 87% and 91% of OUD cases, respectively, among those who have the illness. However, ruling out OUD is also important. Thus, the Drug Abuse Screening Tool (DAST) may be a more appropriate screening option.
For pregnant women with OUD, any treatment improves health-related quality of life (HRQoL), lowers mortality, and decreases adverse birth outcomes such as stillbirth, preterm birth, and NAS.\textsuperscript{17-20} Methadone maintenance treatment (MMT) is used for pregnant women and the general population, and has been shown to have good treatment adherence in clinical trials, with only 22\% of individuals discontinuing treatment.\textsuperscript{21} Buprenorphine maintenance treatment (BMT) is also an option for pregnant women with OUD. Although BMT is more costly than MMT and is associated with lower treatment adherence, pregnant women with OUD treated on BMT have given birth to infants with less severe cases of NAS compared to pregnant women treated with MMT.\textsuperscript{21-24}

Identification and treatment of OUD is possible through effective screening and treatment policies. Clinicians, health plans, health care delivery systems, and policymakers are all relevant stakeholders. From a health sector prospective, the value of screening and treating pregnant women with OUD should incorporate the diagnostic accuracy and cost of screening, the cost of treatment, the cost-savings from adverse events averted through treatment, and the improvements in health-related QOL associated with being illness-free for both women with OUD and infants born to women with OUD. To inform the health sector in their decisions, we developed a decision-analytic model to compare the various screen-and-treat strategies for pregnant women with OUD. This study was approved by the Institutional Review Board at the University of Minnesota.

2.2 Methods

2.2.1 Screening for OUD

We compared three screening tools for OUD – the 4P, SURP, and DAST – to no screening. The tools are comprised of questions aimed at identifying problematic opioid
use. Because the questions vary, each screening tool has a unique ability to identify OUD (sensitivity) and rule out OUD (specificity). The test characteristics of each screening tool are shown in Table 2.1. In our model, women who screened positive for problematic opioid use were referred to a substance use specialist for a definitive diagnosis of OUD prior to treatment. Furthermore, we compared two options for screening frequency, which included screening once during the first prenatal visit versus screening twice at the first and second prenatal visits. In our model, all pregnant women were screened at a starting age of 26, which reflects the median age of pregnancy in the U.S.

2.2.2 Treatment for OUD

In the model, we simulated pregnant women to receive treatment after they screened positive and received a definitive diagnosis of OUD from a specialist. The two treatment options we compared – MMT and BMT – are evidence based pharmacotherapies needed daily to combat the effects of withdrawal. The effectiveness and cost of each medication is shown in Table 2.1. Briefly, being on treatment, regardless of which type, decreased the mortality risks associated with overdose for mothers and significantly decreased stillbirths, preterm births, and infants born with NAS. Infants born to mothers treated with BMT suffer from less severe NAS, resulting in fewer days spent hospitalized, compared to those born to mothers treated with MMT. Receiving either treatment significantly increases HRQoL for mothers; however, the benefits of treatment
<table>
<thead>
<tr>
<th>Variable description</th>
<th>Mean (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cohort characteristics</strong></td>
<td>Value</td>
</tr>
<tr>
<td>Age</td>
<td>26</td>
</tr>
<tr>
<td>Number of pregnancies</td>
<td>1</td>
</tr>
<tr>
<td>Screening frequency at first prenatal visit or at first and second prenatal visit</td>
<td>lifetime</td>
</tr>
<tr>
<td><strong>Screening Tools – beta distribution</strong></td>
<td>Sensitivity</td>
</tr>
<tr>
<td>4P's Plus</td>
<td>0.87 (0.06)</td>
</tr>
<tr>
<td>Drug Abuse Screening Test-10</td>
<td>0.47 (0.06)</td>
</tr>
<tr>
<td>Substance Use Risk Profile-Pregnancy Scale</td>
<td>0.91 (0.04)</td>
</tr>
<tr>
<td><strong>Probabilities – beta distribution</strong></td>
<td>Prevalence of OUD, among pregnant women</td>
</tr>
<tr>
<td></td>
<td>No OUD</td>
</tr>
<tr>
<td></td>
<td>0.005; 0.025</td>
</tr>
<tr>
<td></td>
<td>-</td>
</tr>
<tr>
<td>Preterm birth</td>
<td>0.096 (0.0001)</td>
</tr>
<tr>
<td></td>
<td>0</td>
</tr>
<tr>
<td><strong>Hazard Ratio – lognormal distribution</strong></td>
<td>Hazard Ratio</td>
</tr>
<tr>
<td></td>
<td>Preterm mortality – infant model</td>
</tr>
<tr>
<td></td>
<td>Age 1 – 5</td>
</tr>
<tr>
<td></td>
<td>Age 6 – 12</td>
</tr>
<tr>
<td></td>
<td>Age 13 – 17</td>
</tr>
<tr>
<td></td>
<td>Age 18 – 36</td>
</tr>
<tr>
<td><strong>Costs (2018 dollars) – gamma distribution</strong></td>
<td>Costs</td>
</tr>
<tr>
<td>Screen</td>
<td>$18.90</td>
</tr>
<tr>
<td>Specialist to diagnose OUD - calculated</td>
<td>$119.80</td>
</tr>
<tr>
<td>Time - assumed</td>
<td>1 hour; 2 hours</td>
</tr>
<tr>
<td>Cost</td>
<td>$115.00</td>
</tr>
<tr>
<td>Weekly cost of treatment</td>
<td>$248.82</td>
</tr>
<tr>
<td>Birth</td>
<td>$2,410.77</td>
</tr>
<tr>
<td>Annual healthcare utilization median cost</td>
<td>$994.37</td>
</tr>
<tr>
<td>Age 18 – 44</td>
<td>$5,428.42</td>
</tr>
<tr>
<td>Age 45 – 64</td>
<td>$2,410.77</td>
</tr>
<tr>
<td>Age 65+</td>
<td>$1,384.52</td>
</tr>
<tr>
<td>Neonatal abstinence syndrome – infant model</td>
<td>$1,384.52</td>
</tr>
<tr>
<td>N days in NICU</td>
<td>0</td>
</tr>
<tr>
<td>Cost/day in NICU</td>
<td>$1,277.82</td>
</tr>
<tr>
<td><strong>Utilities – beta distribution</strong></td>
<td>Utilities</td>
</tr>
<tr>
<td>Pregnant/Not pregnant woman</td>
<td>1</td>
</tr>
<tr>
<td>Infant</td>
<td>1</td>
</tr>
</tbody>
</table>
were only realized if women remained on treatment. Clinical trial data show that retention on MMT was much higher than on BMT.\textsuperscript{21}

2.2.3 Markov Models

We developed two Markov models – a maternal and infant model – to project the lifetime costs and health benefits of screening and treating maternal OUD in terms of the impact on the mother and infant. Markov models conceptualize clinical situations into a set of health states and health events. Hypothetical individuals populate the model and transition between health states according to transition probabilities, which vary based on treatment status, current health status, or demographic factors like age, and are derived from the literature. Costs (measured in 2018 US dollars) and benefits (measured in quality-adjusted life-years) are accrued over time. Our maternal model captured the health outcomes, costs, and HRQoL of women with OUD, while our infant model captured costs, and HRQoL of being stillborn, preterm, or having NAS. In our simulation, women only had one pregnancy with one infant. Schematics for both the maternal and infant model are shown in Figure 2.1.

Pregnant women entered the maternal model early on in their pregnancy and either have underlying OUD (Pregnant, OUD\textsuperscript{+}) or do not (Pregnant, OUD\textsuperscript{-}). Following a screening test at a prenatal visit, women with OUD could be identified and treated (i.e., transition to Pregnant, Trt OUD\textsuperscript{+}) if they screened positive and received a definitive diagnosis from a substance use specialist. Women with OUD could also screen negative (false-negative screen). Women without OUD may screen positive for OUD (false-positive screen) but a specialist would rule out disease so they would only incur the cost of the specialist visit and not receive unnecessary treatment.
Figure 2.1 Maternal and Infant Markov Model

OUD– = no opioid use disorder
OUD+ = opioid use disorder
Trt = treated
Rel = relapsed
NAS = neonatal abstinence syndrome
We assume that women on treatment will remain on treatment unless they relapse. We assumed that women did not seek additional treatment after relapsing; therefore, no one was able to restart treatment after discontinuing while pregnant. We assumed that treatment for OUD was needed for life.

After birth, women maintained their OUD status (OUD+, OUD-, treated for OUD, or relapsed OUD). After a positive screen, those on treatment faced a monthly risk of experiencing a relapse. Each month women could die from OUD or from age-related causes.

The infant model consisted of 5 health states, which represent the birth outcomes affected by OUD: being preterm, having NAS, having neither, being both preterm and having NAS, or dead (Figure 2.1). The infant model, like the maternal model, was a lifetime model. Infants could not transition between the birth outcome states. Rather, these health states served as the initial stratification for the lifetime healthcare costs, HRQoL, and mortality that was associated with the outcome. For infants who were stillborn, we included the costs of stillbirth and the loss in life.

2.2.4 Transition Probabilities

We set the prevalence of OUD in pregnancy to 1.5% based on a screening study.\textsuperscript{19} The cumulative rate of developing OUD during pregnancy was determined from clinical input, from Dr Levy, and set to 0.5%. The Maternal Opioid Treatment: Human Experimental Research (MOTHER) clinical trial reported that 48% of pregnant women on BMT and 22% of pregnant women on MMT discontinued treatment.\textsuperscript{21} We assumed that treatment benefits ceased for pregnant women as soon as they relapsed. Newborns received benefits if treatment was taken until at least the beginning of the last trimester.
Stillbirth and preterm births were much lower for women who remained on treatment compared to those who relapsed. Because the relationship between OUD and stillbirth is not well understood, we used the rates of stillbirth reported among marijuana users (2.2 times higher among users than non-users). Preterm birthrates for women with OUD were set to what was reported in a population-based study. Infants born to mothers who were treated with BMT and MMT spent approximately 10 and 17 days in the hospital, respectively. Infants born to mothers who were untreated spent approximately 17 days in the hospital. The additional risk of mortality for someone with OUD was 0.01 times higher than someone without OUD.

2.2.5 Health Sector Perspective

We assumed a health sector perspective. Costs include direct medical costs paid for screening and treatment. We did not include out-of-pocket costs or non-medical costs, including productivity loss. Costs were inflated to 2018 dollars using the CMS Personal Healthcare Expenditure Deflator.

2.2.6 Cost of Screening

We set the cost of screening to $18.90 per individual, regardless of screening test used, and assumed that the prenatal visit costs were the same for those screened and not. The amount of time required for a substance use specialist to diagnose OUD was estimated to be 1 hour, on average, and we varied this time between ½ hour and 2 hours in sensitivity analyses. The cost to see a substance use specialist was set to $119.80 per hour, which is the base salary for a physician who diagnoses and treats injuries reported by the Bureau of Labor Statistics. We assumed that all women who screened positive for OUD had a follow-up with a substance use specialist.
2.2.7 Cost of Treatment

The weekly cost of MMT and BMT was set to $126 and $115, respectively.\textsuperscript{18} For MMT, weekly costs included a daily drug cost of $3, along with a $15 per day cost for integrated psychosocial and medical support services. For BMT, costs included the drug cost of $10 per day and an estimated non-drug cost of $22.50 per visit.\textsuperscript{47} Treatment costs were incurred until treatment stopped.

2.2.7 Health Services Utilization and Cost

Birth costs included any hospitalizations during the nine months prior to the delivery date and all costs incurred during the delivery hospitalization. Costs are reported separately for term and preterm births, and were taken from published administrative claims data reports.\textsuperscript{18} NAS costs were estimated by taking the average days spent with NAS under each treatment and multiplying it by the daily cost of a neonatal intensive care unit stay: $1,278 per day.\textsuperscript{10,21,22}

Mothers without OUD and infants unaffected by OUD were projected to have age-specific medical costs consistent with the general population, which we obtained from the Center for Financing, Access and Cost Trends, Agency for Healthcare Research and Quality: Medical Expenditure Panel Survey.\textsuperscript{34} We multiplied the medical costs of someone without OUD by a factor of 1.4 and 1.28 for women with untreated OUD and treated OUD, respectively. These multipliers were derived from studies that compared the healthcare utilization among those with treated and untreated OUD to those without OUD.\textsuperscript{32,33} The cost of health care utilization for preterm infants were taken from published literature which found that preterm infants required slightly more health care than term infants before the age of 8.\textsuperscript{48}
2.2.8 Health-Related Quality of Life

We found evidence that HRQoL was lower among women with untreated OUD (0.6780) and that HRQoL improved for women on treatment (0.8673).\textsuperscript{36} We found that being preterm decreased HRQoL by 0.092 points.\textsuperscript{41} We assumed this decrease in HRQoL, for preterm infants, until 18 years of age.\textsuperscript{40} We assumed that pregnancy and NAS did not affect HRQoL.

2.2.9 Markov Simulation

We used a weekly time step. Our model accounted for the prevalence of OUD, diagnostic characteristics of screening tools and their costs, outcomes associated with treatment and its associated costs and impact on mortality and HRQoL in the long term. All future costs and quality-adjusted life years (QALYs) were discounted at 3% per year.\textsuperscript{49} We developed and tested the model in TreeAge v. 2019. Some graphics were created using RStudio v. 1.0.136.

2.2.9 Cost-Effectiveness Analysis

For the cost-effectiveness analysis, we followed the guidelines outlined by the 2nd Panel on Cost-Effectiveness in Health and Medicine.\textsuperscript{49} First, strategies were ranked by increasing costs. Dominated strategies (those with higher costs and lower effectiveness than other strategies) were eliminated. For the remaining strategies, we calculated the incremental cost-effectiveness ratio (ICER), defined as the additional cost divided by the additional QALYs gained compared with the next least costly strategy. The best-value, or cost-effective, strategy was identified as the strategy with the highest ICER that was under a threshold of $100,000 per QALY.
2.2.9 Additional Analysis

We calculated the number of adverse outcomes associated with each treatment strategy by tracking the proportion of women that died from OUD, the proportion of stillborn infants, those born with NAS, and those born preterm.

2.2.9 Sensitivity Analysis

We performed 1-way, 2-way, and probabilistic sensitivity analyses (PSA). In the 1-way sensitivity analysis, we varied each parameter, one at a time, between its 95% confidence interval (CI) to determine the impact on the ICER for the cost-effective strategy. For 2-way comparisons, we set our willingness-to-pay (WTP) to $100,000 per QALY and varied two parameters to identify critical values where a specific strategy would always, or never, be cost-effective. In the PSA, we calculated ICER point estimates using a random sample of 10,000 sets of parameters from their estimated probability distributions (specified in Table 2.1). We used these model outputs to construct the cost-effectiveness acceptability curve (CEAC) which displays the proportion of times a strategy was cost-effective among the 10,000 samples for varying WTP thresholds.

2.3 Results

2.3.1 Cost-Effectiveness Analysis

In all, we compared 13 screen-and-treat strategies. The costs and benefits of each strategy are reported for the maternal and infant model separately and combined in Table 2.2; however, we report the combined results in text. The least effective strategy was the no screen, no treat strategy, which was associated with a lifetime cost of $401,023 and lifetime benefit of 55.92 QALYs. Most strategies were dominated (higher costs and
<table>
<thead>
<tr>
<th>Maternal Model</th>
<th>Cost</th>
<th>QALY</th>
<th>Infant Model</th>
<th>Cost</th>
<th>QALY</th>
<th>Combined Model</th>
<th>Cost</th>
<th>QALY</th>
<th>ICER</th>
</tr>
</thead>
<tbody>
<tr>
<td>No screen, no treatment</td>
<td>$160,976</td>
<td>26.5346</td>
<td>$240,047</td>
<td>29.3809</td>
<td>$401,023</td>
<td>55.9155</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4P once + BMT</td>
<td>$161,204</td>
<td>26.5395</td>
<td>$239,881</td>
<td>29.3872</td>
<td>$401,085</td>
<td>55.9267</td>
<td>$5,494</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4P once + MMT</td>
<td>$161,463</td>
<td>26.5466</td>
<td>$239,876</td>
<td>29.3889</td>
<td>$401,339</td>
<td>55.9350</td>
<td>$28,493</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4P twice + MMT</td>
<td>$161,544</td>
<td>26.5478</td>
<td>$239,854</td>
<td>29.3898</td>
<td>$401,399</td>
<td>55.9377</td>
<td>$28,764</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SURP twice + MMT</td>
<td>$161,581</td>
<td>26.5480</td>
<td>$239,853</td>
<td>29.3899</td>
<td>$401,434</td>
<td>55.9380</td>
<td>$122,945</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAST once + BMT</td>
<td>$161,1150</td>
<td>26.5371</td>
<td>$239,960</td>
<td>29.3842</td>
<td>$401,075</td>
<td>55.9214</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAST twice + BMT</td>
<td>$161,187</td>
<td>26.5382</td>
<td>$239,911</td>
<td>29.3858</td>
<td>$401,098</td>
<td>55.9241</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SURP once + BMT</td>
<td>$161,234</td>
<td>26.5397</td>
<td>$239,874</td>
<td>29.3874</td>
<td>$401,108</td>
<td>55.9271</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4P twice + BMT</td>
<td>$161,260</td>
<td>26.5399</td>
<td>$239,860</td>
<td>29.3879</td>
<td>$401,120</td>
<td>55.9278</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SURP twice + BMT</td>
<td>$161,292</td>
<td>26.5400</td>
<td>$239,859</td>
<td>29.3879</td>
<td>$401,151</td>
<td>55.9280</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAST once + MMT</td>
<td>$161,252</td>
<td>26.5409</td>
<td>$239,957</td>
<td>29.3851</td>
<td>$401,209</td>
<td>55.9261</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAST twice + MMT</td>
<td>$161,380</td>
<td>26.5436</td>
<td>$239,907</td>
<td>29.3872</td>
<td>$401,287</td>
<td>55.9309</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SURP once + MMT</td>
<td>$161,504</td>
<td>26.5471</td>
<td>$239,869</td>
<td>29.3893</td>
<td>$401,374</td>
<td>55.9365</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4P = 4Ps Plus  
SURP = Substance Use Risk Profile  
DAST = Drug Abuse Screening Tool  
BMT = buprenorphine maintenance treatment  
MMT = methadone maintenance treatment  
NAS = neonatal abstinence syndrome  
QALY = quality-adjusted life year  
ICER = incremental cost-effectiveness ratio

Once = screen at first prenatal visit  
Twice = screen at first and second prenatal visit
lower benefits than another strategy), especially the screening strategies that involved screening with DAST and treatment with BMT.

In cost-effectiveness analyses, the strategy that is cost-effective depends on ones WTP. For a WTP of $100,000 per QALY, the cost-effective strategy was to screen at the first and second prenatal visit with the 4P and treat those who screen positive with MMT (4P twice + MMT). The 4P twice + MMT strategy was associated with a lifetime cost of $401,399 and lifetime benefit of 55.94 QALYs, which amounts to an ICER of $28,764.

2.3.2 Additional Analysis

We found that doing nothing (no screen, no treat strategy) would result in 20 maternal overdose deaths, 103 stillbirths, 1,659 preterm births, and 6,693 infants born with NAS among 10,000 pregnant women. Implementing the cost-effective strategy (4P twice + MMT) would reduce the number of opioid related maternal deaths, stillbirths, preterm births, and NAS by 90, 36.9, 26.5, and 45.4 percent, respectively (Table 2.3).

2.3.3 Sensitivity Analysis

In the 1-way sensitivity analysis (Figure 3), we found that our results were most sensitive to the sensitivity of the 4P screening tool as the ICER changed exponentially when we varied this parameter between its 95% CI. Specifically, when the sensitivity of the 4P improved, the 4P once + MMT strategy became the cost-effective strategy. This is because a better performing screening tool would eliminate the need to screen more than once. Another parameter that affected the cost-effectiveness of the 4P twice + MMT strategy was the rate at which patients relapsed on MMT. Our analysis showed that if relapse on MMT increased, then the 4P twice + MMT strategy’s ICER increased. This makes sense because when individuals relapsed in the model, they no longer received the
Table 2.3 Adverse Outcomes Per 10,000 OUD-Related Pregnancies after 40 weeks, N(% reduction)

<table>
<thead>
<tr>
<th></th>
<th>No screen, no treat</th>
<th>Screen-and-Treat Strategies</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>4P once + BMT</td>
</tr>
<tr>
<td><strong>Maternal deaths</strong></td>
<td>20</td>
<td>6 (70.0)</td>
</tr>
<tr>
<td><strong>Stillbirth</strong></td>
<td>103</td>
<td>74 (28.2)</td>
</tr>
<tr>
<td><strong>Preterm birth</strong></td>
<td>1,659</td>
<td>1,372 (17.3)</td>
</tr>
<tr>
<td><strong>NAS</strong></td>
<td>6,693</td>
<td>4,002 (40.2)</td>
</tr>
</tbody>
</table>

4P = 4Ps Plus
SURP = Substance Use Risk Profile
BMT = buprenorphine maintenance treatment

Once = screen at first prenatal visit
Twice = screen at first and second prenatal visit
MMT = methadone maintenance treatment
NAS = neonatal abstinence syndrome
Figure 2.2 1-Way Sensitivity Analysis (4P twice + MMT vs 4P once + MMT)

Incremental cost-effectiveness ratio (ICER)
- ICER at $100,000 per quality-adjusted life year
- Base-case ICER of screen with 4P twice and treat with methadone maintenance strategy (4P twice + MMT)
benefits of treatment. Though it is worth noting that despite increasing the relapse rate to its highest bound CI, the 4P twice + MMT strategy was still cost-effective as the ICER never increased to > $100,000 per QALY. The next parameter that had some impact on the 4P twice + MMT strategy was the cost of a definitive screen with a substance use specialist. When screening got more expensive, the 4P twice + MMT strategy’s ICER increased. However, even at the highest value of the parameter’s CI, the 4P twice + MMT strategy remained cost-effective at the threshold value of $100,000 per QALY gain. Most other parameters had minimal effect on the 4P twice + MMT strategy, as the ICER remained quite steady across all ranges (see probability of NAS on MMT as an example).

In the 2-way sensitivity analysis, we found the threshold value for when the SURP could overtake the 4P as the cost-effective screening strategy. In our base case, the 4P had the ability to identify 87% of OUD cases. Our analysis showed that if the performance of the 4P fell to 86%, then the SURP would be the better-value screening option. Furthermore, we found the threshold value for when BMT could overtake MMT as the cost-effective strategy. In our base case, the relapse rate on BMT was 0.48. Our analysis showed that if relapse on BMT improved to 0.3 then BMT would be the better-value treatment option.

The PSA showed that the no screen, no treat strategy was cost-effective at WTP < $15,000. Between a WTP of $15,000 and $31,000, the cost-effective strategy was to screen once with the 4P and treat with BMT. Between a WTP of $31,000 and $39,000 the cost-effective strategy was to screen once with the 4P and treat with MMT. The 4P twice + MMT strategy became the cost-effective strategy at WTP > $39,000 and < $115,000 (Figure 5). Among 10,000 simulations, the 4P twice + MMT strategy would have a 45%
Figure 2.3 2-Way Sensitivity Analysis

- **Base-case**
- MMT = methadone maintenance treatment
- BMT = buprenorphine maintenance treatment

4P = 4Ps Plus
- SURP = Substance Use Risk Profile
- Once = screen at first prenatal visit
- Twice = screen at first and second prenatal visit
Figure 2.4 Cost-Effectiveness Acceptability Curve

4P = 4Ps Plus
SURP = Substance Use Risk Profile
Once = screen at first prenatal visit
Twice = screen at first and second prenatal visit
MMT = methadone maintenance treatment
BMT = buprenorphine maintenance treatment

- No screen + no treat
- 4P once + BMT
- 4P once + MMT
- 4P twice + MMT
- SURP twice + MMT
- All other dominated strategies

Cost-effective strategy
chance of costing less than $100,000. For WTP > $115,000 the cost-effectiveness strategy was to screen twice with the SURP and treat with MMT.

2.4 Discussion

Screening and treating OUD is cost-effective for pregnant women. This cost-effective analysis identified a best-value screen-and-treat strategy for pregnant women with OUD. Screening at the first and second prenatal visit with the 4P and treating with MMT was cost-effective at WTP values between $39,000 and $115,000 per QALY. Furthermore, screening and treating OUD would decrease maternal morbidity, and adverse birth outcomes such as stillbirths, preterm births, and infants born with NAS. Our findings support clinical guidelines that recommend universal screening for OUD. Policymakers and clinicians should consider setting the 4P as the recommended screening tool. Furthermore, MMT should be offered to all women with OUD.

Prior cost-effectiveness studies of MMT vs BMT have found mixed results. Our study produced similar finding to a past cost-effective analysis that compared MMT to no therapy for OUD in a general population. In that study, MMT had an ICER of $5,915 per life-year gained compared to no intervention. In another review, MMT was slightly more effective and less costly than BMT (QALY difference ranging from 0.00055 to 0.0126). In 2019, a very similar study comparing MMT, BMT, and detoxification during pregnancy for OUD was published. Researchers found that BMT was the most cost-effective strategy but did not include screening in their model, and used different parameter estimates which included relapse rates, rates of NAS, and cost inputs. To the best of our knowledge, our study was the first to include screening for OUD with treatment. Studies that have looked at screening for a combination of substance use
disorders do exist. A review found that there were cost-savings from screening and brief intervention for alcohol use disorder.\textsuperscript{53} While another study found that screening, and brief intervention for multiple substance use disorders in an Emergency Department setting did not have any significant impact on society.\textsuperscript{54} We suspect that a lot of the differences found can be attributed to study design, the population studied, and the different outcomes measured.

\textit{2.4.1 Societal Perspective}

We assumed a healthcare sector perspective because this closely mirrors the perspective of policymakers. However, a societal perspective could also be useful to policymakers because policy would affect not just those who have OUD but society in general. From a societal perspective, all costs, regardless of who pays for them, would be included in the model. Therefore, in addition to the cost of the intervention itself and downstream healthcare cost savings, the cost of transportation to get treatment, time lost at work due to illness or to receive care, cost of childcare while receiving care, cost to the criminal justice and drug enforcement system, and social services would also be included. We suspect that taking on a societal perspective would likely increase the cost-effectiveness of screening for and treating OUD.

In addition, for any additional years of life that would be gained by screening and treating OUD, the 2nd Panel on Cost-effectiveness in Health and Medicine recommends that all consumption costs incurred or earnings accumulated be included in the analysis.\textsuperscript{49} It is not clear whether these additional survivor costs would increase or reduce the ICERs; therefore, it is not clear how including these societal factors would affect the ICERs. Currently, there is a debate regarding whether survivor consumption costs and earnings
should be included. Because of the unsettled nature of the recommendations, we have opted not to include a societal perspective analysis in this paper. Policymakers should be aware of how a societal perspective may change what is deemed the best-value strategy.

2.4.1 Limitations

One limitation of this study is that it understates the full benefits of identifying and treating OUD because of polysubstance abuse disorder, which we did not model, but should be considered in future analyses. Other studies exist that have taken on a more societal perspective of addiction to heroin and prescription drugs and their findings warrant attention. Briefly, one study found that the societal burden of heroin addiction in 2015 was approximately $51.2 billion. These costs included not just healthcare utilization, but also cost of overdose deaths, loss of productivity due to incarceration, crime costs associated with heroin use, and NAS. Moreover, another study found that the total societal costs of prescription opioid abuse was $25.6 billion from lost workplace productivity, $25.0 billion from healthcare utilization, and $5.1 billion from criminal justice costs.

Pregnant women with OUD may have mental health conditions, such as depression, history of trauma, posttraumatic stress disorder, and anxiety. According to a recent study, among a sample of 125 postpartum women seeking treatment for OUD, at least 30% also needed treatment for depression, and at least 40% showed symptoms of postpartum depression. Individuals with OUD may also use substances such as tobacco, marijuana, and cocaine. Tobacco use in pregnancy has been well studied and has been linked to very poor birth outcomes. None of these additional substance use disorders
were represented in our model but we can assume that screening and treating OUD could facilitate treatment for other conditions.

The main limitation of the model is that we have simplified OUD over the course of a lifetime into just a few health states that may not appropriately reflect all possible health scenarios. For example, preterm infants often suffer from other serious conditions that were not accounted for in our analysis, pregnant women may experience birth complications that we did not include, and OUD may have differing levels of dependence so that some women may not need treatment for the entirety of their lives. Furthermore, treatment is available even after relapsing but we did not allow those who relapsed to reenter treatment. In such situations, our findings would be an underestimate of the true benefits of screening and treatment.

2.4.2 Conclusion

Screening is only beneficial if treatment is available; therefore, clinicians should only screen for OUD if they are prepared to offer treatment. We found that the 4P was the best-value screening tool; however, in situations where the 4P is not available, then clinicians should screen with any tool because we found that screening and treatment was always better than the “do nothing” strategy. Our model showed that screening more than once was cost-effective. This is because the benefits associated with OUD detection and treatment were so great that they outweighed the cost of a second screen. We know that screening can be burdensome for some providers; however, if providers are able to screen at least once, they can achieve approximately 88% of the benefit achievable by screening twice. We found that MMT was the best-value treatment; however, if MMT is not an
option because of patient preferences or other barriers that may prevent access, then BMT should be offered.
Chapter 3 Future Research on Treatment for Opioid Use Disorder in Pregnancy: A Value of Information Analysis

3.1 Introduction

Despite ongoing efforts to combat the opioid epidemic in the general population, less attention has been paid to pregnant women, who also suffer from opioid use disorder (OUD). OUD in pregnancy not only increases the mother’s morbidity and mortality but also contributes to serious birth outcomes including stillbirth, preterm birth, and neonatal abstinence syndrome (NAS). Multiple screening tools and treatment options are available to identify and treat women with OUD in pregnancy.

In the previous chapter, I constructed a Markov model to estimate the benefits and costs of various screen-and-treat strategies for OUD in a cohort of 26-year-old pregnant women. I found that screening pregnant women for OUD at their first and second prenatal visit with the 4Ps Plus (4P) was cost-effective when compared to the Substance-Use Risk Profile (SURP) and the Drug Abuse Screening Tool (DAST). Although the SURP had a higher sensitivity (0.91 v 0.87) and the DAST had a higher specificity (0.82 v 0.76) compared to the 4P, the prevalence of OUD in pregnancy justified the 4P as the better-value screening tool. Following screening, methadone maintenance treatment (MMT) was recommended over buprenorphine maintenance treatment (BMT). Although BMT was associated with a lower prevalence and severity of NAS in infants, its high costs and high relapse rates, weakened its benefits when compared to MMT.

However, like many cost-effective studies, my findings are highly dependent on the model’s inputs and assumptions. This was shown to be the case in sensitivity analyses (SA), where I found that if the true value of the 4P’ sensitivity were 0.86 (down from
0.87) and all other parameters were correctly estimated, then the cost-effective screening tool would switch from the 4P to the SURP. In addition, SA showed that if the true value of relapse on BMT was more similar to MMT (0.48 to 0.22), then the cost-effective treatment would switch from MMT to BMT. Although current information suggests that screening with the 4P followed by MMT for those who test positive is the best-value strategy, there is a possibility that my model has identified a suboptimal strategy, due to model assumptions, which would result in reduced net health benefits.

Given that there is uncertainty in our decision, it may be beneficial to collect more information (i.e., conduct additional research) to reduce the uncertainty in the parameter values and increase the chance that the optimal strategy is identified (and thus maximize net health benefits). However, additional research is costly, and ultimately may not change the recommendation for the best-value treatment. A formal analysis, called a value of information (VOI) analysis, can help identify if additional research is warranted.

Additional Research Ideas

First, I consider a potential observational study to re-evaluate the performance of the screening tools for OUD during pregnancy. This new study has the ability to reduce the current uncertainty surrounding the existing test characteristics, which I assumed in my decision model to be the same, regardless of when screening occurred or how many times an individual was screened.

Another observational study I consider is one that could collect more information on the quality of life (QOL) of those treated with MMT and BMT. In my model, I assumed that QOL was the same for MMT and BMT, and that QOL was only affected by
being on or off treatment. The existing study that evaluated QOL of treatment enrolled just 44 individuals. This may be too small a sample size. An additional study with a larger sample size might eliminate some of the uncertainty surrounding QOL and improve our chances of identifying the optimal screen-and-treat strategy.

The final potential study I consider is a randomized control trial (RCT). The trial would follow pregnant women on MMT or BMT. It would track the following: 1) relapse on treatment; 2) timing of relapse; and 3) maternal and infant outcomes. This study is needed because the timing of relapse during pregnancy may lead to different outcomes for mothers and infants. I was unable to find literature on how the timing of relapse affected maternal and infant outcomes. I made an assumption in my decision model that if relapse occurred in the first two trimesters of pregnancy, then both the mother and infant would lose all benefits associated with treatment, and that for those who relapsed in the third trimester, all infant benefits were still gained; an additional study would relax these assumptions, and provide better parameter estimates of relapse, and treated vs. untreated outcomes.

3.2 Methods

In this chapter, I conducted a VOI analysis. I tested two different study designs (the RCT and observational studies described previously) on three different parameter sets (screening tools performance; outcomes associated with timing of relapse and treatment; and quality of life associated with treatment) to determine whether conducting more research is warranted.
3.2.1 Markov Model Review

The Markov model I constructed in the previous chapter compared 12 different screen-and-treat strategies as well as no screening or treatment. The screen-and-treat strategies varied by number of screens (first visit or first and second visit), type of screen (4P, SURP, or DAST), and treatment for OUD (MMT or BMT).

The model projected the lifetime costs and benefits, measured in quality-adjusted life years (QALYs), for each strategy among a cohort of pregnant women with and without OUD. Pregnant women with OUD could either screen positive or negative; those who screened positive were referred a specialist for confirmatory screening and then to treatment. They incurred the costs and benefits of treatment until they relapsed or died. Infants born to women with OUD who were on treatment were less likely to be born preterm or have NAS. Alternately, pregnant women (and their child) who had OUD but were not identified were more likely to be born preterm and have NAS. Pregnant women who did not have OUD only incurred the cost of screening, and additionally, the cost to see a substance use specialist if they were mistakenly referred to treatment (as a result of a false-positive screen).

Table 3.1 contains the model inputs and the distributions used to reflect uncertainty. Screening test performance was taken from randomized controlled trials and treatment efficacy was derived from the MOTHER clinical trial.\textsuperscript{15,16,21,60} Clinical expertise was used in estimating the time needed to diagnose OUD following a positive screening test and the prevalence of OUD in pregnancy. Following the recommendations
<table>
<thead>
<tr>
<th>Variable description</th>
<th>Mean (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cohort characteristics</strong></td>
<td><strong>Value</strong></td>
</tr>
<tr>
<td>Age</td>
<td>26</td>
</tr>
<tr>
<td>Number of pregnancies</td>
<td>1</td>
</tr>
<tr>
<td>Screening frequency</td>
<td>at first prenatal visit or at first and second prenatal visit</td>
</tr>
<tr>
<td>Duration of treatment</td>
<td>lifetime</td>
</tr>
<tr>
<td>4P’s Plus²⁶</td>
<td>0.87 (0.06)</td>
</tr>
<tr>
<td>Drug Abuse Screening Test-10²⁵</td>
<td>0.47 (0.06)</td>
</tr>
<tr>
<td>Substance Use Risk Profile-Pregnancy Scale²⁷</td>
<td>0.91 (0.04)</td>
</tr>
<tr>
<td><strong>Screening Tools – beta distribution</strong></td>
<td><strong>Sensitivity</strong></td>
</tr>
<tr>
<td>4P’s Plus²⁶</td>
<td>0.76 (0.03)</td>
</tr>
<tr>
<td>Drug Abuse Screening Test-10²⁵</td>
<td>0.82 (0.03)</td>
</tr>
<tr>
<td>Substance Use Risk Profile-Pregnancy Scale²⁷</td>
<td>0.65 (0.07)</td>
</tr>
<tr>
<td><strong>Probabilities – beta distribution</strong></td>
<td><strong>No OUD</strong></td>
</tr>
<tr>
<td>Prevalence of OUD, among pregnant women²⁸</td>
<td>-</td>
</tr>
<tr>
<td>Develop OUD²³</td>
<td>0.005; 0.025</td>
</tr>
<tr>
<td>Relapse²⁹</td>
<td>-</td>
</tr>
<tr>
<td>OUD related mortality²⁸,³⁰</td>
<td>0.48 (0.07)</td>
</tr>
<tr>
<td>Stillbirth²⁰</td>
<td>-</td>
</tr>
<tr>
<td>Preterm birth¹⁷</td>
<td>0.17 (0.02)</td>
</tr>
<tr>
<td>Neonatal abstinence syndrome – infant model¹⁸,³⁹</td>
<td>-</td>
</tr>
<tr>
<td><strong>Hazard Ratio – lognormal distribution</strong></td>
<td><strong>Pre-term infant</strong></td>
</tr>
<tr>
<td>Age 1 – 5</td>
<td>1.53 (0.2992)</td>
</tr>
<tr>
<td>Age 6 – 12</td>
<td>1.18 (0.2321)</td>
</tr>
<tr>
<td>Age 13 – 17</td>
<td>1.28 (0.2219)</td>
</tr>
<tr>
<td>Age 18 - 36</td>
<td>1.31 (0.0944)</td>
</tr>
<tr>
<td><strong>Costs (2018 dollars) – gamma distribution</strong></td>
<td><strong>No OUD</strong></td>
</tr>
<tr>
<td>Screen²²</td>
<td>$18.90</td>
</tr>
<tr>
<td>Specialist to diagnose OUD - calculated</td>
<td>-</td>
</tr>
<tr>
<td>Time - assumed</td>
<td>1 hour; 2 hours</td>
</tr>
<tr>
<td>Cost²³</td>
<td>$119.80</td>
</tr>
<tr>
<td>Weekly cost of treatment²⁴</td>
<td>-</td>
</tr>
<tr>
<td>Birth²²</td>
<td>-</td>
</tr>
<tr>
<td>Full-term birth</td>
<td>$18,253.83</td>
</tr>
<tr>
<td>Preterm birth</td>
<td>$45,605.58</td>
</tr>
<tr>
<td>Annual healthcare utilization median cost¹⁰,¹¹,³⁶</td>
<td>$944.37</td>
</tr>
<tr>
<td>Age 18 – 44</td>
<td>$2,410.77</td>
</tr>
<tr>
<td>Age 45 – 64</td>
<td>$5,428.42</td>
</tr>
<tr>
<td>Age 65+</td>
<td>$904.37</td>
</tr>
<tr>
<td>Neonatal abstinence syndrome – infant model</td>
<td>-</td>
</tr>
<tr>
<td>N days in NICU¹³,¹⁵</td>
<td>0.015</td>
</tr>
<tr>
<td>Cost/day in NICU³⁸</td>
<td>$1,777.82</td>
</tr>
<tr>
<td><strong>Costs (2018 dollars) – gamma distribution</strong></td>
<td><strong>Term infant</strong></td>
</tr>
<tr>
<td>Age 1</td>
<td>$4,986.54</td>
</tr>
<tr>
<td>Age 2</td>
<td>$1,991.62</td>
</tr>
<tr>
<td>Age 3 – 4</td>
<td>$991.31</td>
</tr>
<tr>
<td>Age 5 – 7</td>
<td>$706.36</td>
</tr>
<tr>
<td>Age 8 – 17</td>
<td>$2,969.43</td>
</tr>
<tr>
<td>Utilities – beta distribution</td>
<td><strong>No OUD</strong></td>
</tr>
<tr>
<td>Pregnant/Not pregnant woman³⁰</td>
<td>1</td>
</tr>
<tr>
<td>Utilities – beta distribution</td>
<td><strong>Term infant</strong></td>
</tr>
<tr>
<td>Infant³¹,³²</td>
<td>1</td>
</tr>
</tbody>
</table>
outlined by the 2nd Panel on Cost-Effectiveness in Health and Medicine, probability and utility parameters were assigned beta distributions, and costs were assigned gamma distributions. All parameters were assumed to be independent.

### 3.2.2 Probabilistic Sensitivity Analysis

The first step in a VOI is to conduct probabilistic sensitivity analysis (PSA). For the PSA I sampled values for all parameters from their distributions and calculated the expected costs and benefits using the sampled set of parameters, and did this 10,000 times. For each PSA iteration, I determined which strategy was cost-effective, then I calculated the incremental cost-effectiveness ratio (ICER) at various willingness-to-pay (WTP) thresholds.

### 3.2.3 Expected Value of Perfect Information

Next, I calculated the expected value of perfect information (EVPI). EVPI is the maximum amount a decision maker should be willing to pay to eliminate uncertainty from all parameters. To calculate the EVPI, I used the PSA dataset and calculated the net monetary benefit (NMB) of each strategy over each simulation for a range of WTP values (NMB = QALY * WTP – cost). Then, I identified the optimal strategy from each simulation, defined as the strategy with the highest NMB. I also found the overall optimal strategy, which was the strategy with the highest average net benefit across all 10,000 simulations (i.e., the optimal strategy without perfect information). Then, I calculated the difference between the NMB of the optimal strategy identified per PSA iteration and the NMB for the overall optimal strategy per iteration; this difference is referred to as the opportunity loss (if the optimal strategy for the iteration is the same as the overall optimal strategy then the opportunity loss is $0). The opportunity loss represents the cost, or the
negative effect, of selecting a less cost-effective strategy for a given PSA sample. I averaged the opportunity losses over all the 10,000 PSA samples to get the EVPI. I extrapolated EVPI from an individual-level cost to a population-level cost by multiplying it by 6,200,00 which represents approximately the number of pregnant women annually who would be screened for OUD.\textsuperscript{43}

3.2.4 Expected Value of Partial Perfect Information

To calculate the expected value of partial perfect information (EVPPI), I used the linear meta-modeling approach with a Gaussian approximation described by Jalal and Alarid-Escudero and applied in a recent Jutkowitz et al. paper.\textsuperscript{61,62} Briefly, this approach regresses each parameter, or set of parameters, on the opportunity loss (found in previous step) to determine the relationship between the inputs and expected benefits. This method is a more efficient method for calculating VOI, which previously required substantial computational power to resample and calculate outcomes under differing parameters. As with EVPI, EVPPI was extrapolated to a population of 6,200,000. The linear meta-modeling approach with a Gaussian approximation was applied using R Dampack, which was developed by the DARTH working group at the University of Minnesota, Twin-Cities.\textsuperscript{63}

I considered three distinct research studies that could be conducted to collect additional information on subsets of model parameters. The first study was an observational study to evaluate the sensitivity and specificity of the three screening tools for identifying OUD in pregnancy. The second study was an observational study used to inform the health utility values to be assigned to pregnant women with OUD who are treated with MMT and BMT. The last study was a randomized clinical trial (RCT) to
evaluate the efficacy of treatment with timing of relapse. Specific outcomes to be measured in the RCT included relapse on treatment and its effect on maternal and neonatal outcomes, such as mortality, stillbirth, preterm birth, and NAS.

3.2.5 Expected Value of Sample Information and Expected Net Benefit of Sampling

EVSI is a measure of the value of reducing parameter uncertainty by conducting further studies but not reducing the uncertainty completely. In practice, EVSI can be used to determine the optimal sample size for a study that is designed to reduce the uncertainty of a subset of parameters. This is determined by taking into account the cost of research. To determine whether the value of conducting further studies outweighs the cost of conducting those studies, we calculate what is called the expected net benefit of sampling (ENBS). In other words, the ENBS is the difference between the cost of research and EVSI for a given study sample size. The sample size that produces the maximum ENBS (i.e. the optimal sample size) is the number of participants we should not exceed when conducting additional research.

For this analysis, calculating EVSI and ENBS meant that I needed an estimate for the cost of research. I relied on estimates from the Jutkowitz et. al, paper which estimated that the cost of a two-arm randomized controlled trial had a fixed cost of $8.74 million and a per patient cost of $8,440; and that the fixed cost of conducting an observational study was $50,000 with an additional per patient cost of $500.33. I calculated the EVSI and ENBS using the same methods described by Jalal and Alarid-Escudero.61
3.3 Results

3.3.1 Probabilistic Sensitivity Analysis

To display the PSA results, I plotted the difference in cost and effectiveness (incremental cost-effectiveness ratio or ICER) between the SURP and MMT intervention and the 4P and MMT intervention (Figure 3.1). In the figure, each dot represents a possible real-world situation of costs incurred and benefits received. The dotted line represents the WTP threshold. All dots falling below the dotted line were considered cost-effective. All dots falling within the ellipse were within 95% of the average ICER.

The reason for comparing the SURP and MMT intervention to the 4P and MMT intervention is that they were the two most cost-effective strategies. Among 10,000 simulations, the 4P and MMT strategy was cost-effective 45% of the time, and the SURP and MMT strategy which was cost-effective 34% of the time, at a WTP of $100,000 per QALY (Figure 3.2). This means that there was a lot of uncertainty in the cost-effective strategy because no strategy was cost-effective the majority of the time. These findings support the need to conduct a VOI analysis.
Figure 3.1 Incremental Cost-Effectiveness Ratio of SURP + MMT v. 4P + MMT
Figure 3.2 Cost-Effectiveness Acceptability Curve

4P = 4Ps Plus
SURP = Substance Use Risk Profile
Once = screen at first prenatal visit
Twice = screen at first and second prenatal visit
MMT = methadone maintenance treatment
BMT = buprenorphine maintenance treatment

Cost-effective strategy
3.3.2 Expected Value of Population Perfect Information

The population EVPI peaked at WTP values which corresponded to when the cost-effective strategy changed. For example, at a WTP of $20,000 the cost-effective strategy changed from no screen and no treat, to screen with the 4P and treat with BMT (Figure 3.2). To reflect this change, in Figure 3.3 we see our first peak in EVPI (around the same WTP of $20,000). The second peak in EVPI is at approximately $35,000 where the cost-effective strategy changed from screening once with the 4P and treating with MMT, to screening twice with the 4P and treating with MMT. The last peak in EVPI occurred around $115,000 where the cost-effective strategy changed from 4P and MMT, to SURP and MMT. The peaks indicate that there is greater value in reducing uncertainty at the corresponding WTP thresholds.

3.3.3 Expected Value of Population Partial-Perfect Information

The population EVPPI was highest at a WTP of $20,000 and $35,000 for the study designed to estimate the utility values of treatment (Figure 3.4). Furthermore, EVPPI was highest for the study designed to evaluate the performance of the screening tools at a WTP of $40,000 and $115,000 (Figure 3.5). Although, EVPPI peaked for the study designed to estimate the outcomes associated with treatment; it was much lower when compared to the other two study designs (Figure 3.6).
Figure 3.3 Expected Value of Population Perfect Information

Figure 3.4 Expected Value of Population Partial-Perfect Information on Screening with the 4Ps Plus, Substance Use Risk Profile, and Drug Abuse Screening Tool
Figure 3.5 Expected Value of Population Partial-Perfect Information on Utility on Methadone and Buprenorphine Maintenance Treatment

Figure 3.6 Expected Value of Population Partial-Perfect Information on Maternal and Neonatal Outcomes Associated With Timing of Relapse on and off Treatment
3.3.4 Expected Value of Sample Information for Parameters (EVSI) and Expected Net Benefit of Sampling (ENBS)

Figure 3.7 displays the ENBS (blue line) for the study designed to evaluate the performance of the OUD screening tools, at a WTP of $100,000. The red line is the value gained (EVSI) per additional participant (labeled on the x-axis) and the green line represents the cost of enrolling an additional participant. At $100,000, the EVSI is higher than the cost of research. The sample size that would maximize the difference between cost of research and EVSI was 2,400.

Similarly, figure 3.9 displays the ENBS for the study designed to evaluate the utility associated with being on and off treatment, at a WTP of $100,000. The EVSI is higher than the cost of research. The sample size that maximized the difference between cost of research and EVSI was 220.

For the EVSI and ENBS analysis, the RCT needed to evaluate the outcomes associated with relapse and maternal and child outcomes, was not warranted. The cost of conducting an RCT (fixed cost $8.74 million) was too costly to justify the value the study would bring (Figure 3.9). EVSI for the RCT remains very low until a larger sample size is collected; this is illustrated in figure 3.10. EVSI started to increase around a sample size of 200, and at a sample size of 4,000 the EVSI ($735,000) would still not outweigh the fixed cost of implementing an RCT ($8.74 million).
Figure 3.7 Expected Net Benefit of Sampling for Screening with the 4Ps Plus, Substance Use Risk Profile, and Drug Abuse Screening Tool at WTP $100,000

Figure 3.8 Expected Net Benefit of Utility on Methadone and Buprenorphine Maintenance Treatment at WTP $100,000

EVSI = expected value of sample information
COR = cost of research
ENBS = expected net benefit of sampling
OSS = optimal sample size
Figure 3.9 Expected Net Benefit of Sampling for Maternal and Neonatal Outcomes Associated With Timing of Relapse on and off Treatment at WTP $100,000

Figure 3.10 Expected Value of Sample Information for Maternal and Neonatal Outcomes Associated with Timing of Relapse on and off Treatment

EVSI = expected value of sample information
COR = cost of research
ENBS = expected net benefit of sampling
3.4 Discussion

I conducted a VOI analysis to determine the value of conducting additional research to reduce decision uncertainty on the cost-effectiveness of screening and treating maternal OUD. I found that conducting additional research was warranted, if it examined screening performance and the utilities of being on treatment. I used the ENBS to determine the maximum amount of participants this study should enroll. I found that enrolling no more than 2,400 and 220 participants in the observational study examining screening performance and treatment utilities would provide the best-value. Other research study designs, including an RCT focused on treatment relapse and outcomes, was found to be too costly to implement, given that the value it could provide would not exceed the research costs. Future researchers could consider these thresholds when designing future research protocols but should also consider whether the sample sizes indicated would produce meaningful statistical differences.

3.4.1 Limitations

This study has several limitations. First, some of the concepts of VOI, such as EVPPI, EVPI, EVSI and ENBS may be difficult to understand. Fortunately there are many resources that exist which can help; including, a recent publication by ISPOR on conducting and interpreting VOI studies. Next, when deciding if there is value in collecting additional information, the budget, or WTP, of the funding organization needs to be clearly defined. In the U.S. there is no formal threshold on which to base VOI. Therefore, when determining if additional research is needed, individuals or funding organizations must set their own budget, which may require additional thought. Finally, this VOI analysis used current information to determine if collecting additional
information was valuable. This means that the VOI model relied heavily on the parameters, and their known uncertainty (standard errors) displayed in Table 3.1. A thorough literature review was conducted in 2018 to collect that information and it is possible that newer research was missed.

3.4.2 Conclusion

This study used a linear regression meta-modeling approach with a Gaussian approximation to conduct a VOI. The specific parameters that needed to be reevaluated include the performance of the various screening tools for detecting OUD in pregnancy. They include the 4P, SURP, and DAST. Other parameters requiring reevaluation included the utilities attributed to being on MMT and BMT, and a measure of how timing of relapse while on treatment affects maternal and infant outcomes. We found that there was value in waiting, to collect additional information, specifically related to treatment utilities, before making a decision on a best-value screen-and-treat strategy for OUD in pregnancy. The cost of research to conduct studies on screening performance and treatment relapse and consequences were too large to provide value, with a WTP of $100,000.
Chapter 4 Cost Estimators for Censored Data and Their Impact on Cost-Effectiveness Analyses

4.1 Introduction

Cost-effectiveness analyses are used to produce evidence that policymakers may consider when making decisions about health care benefits to optimize value. Because analyses are only as good as the parameters informing the model, obtaining non-biased parameter inputs are just as important as conducting the cost-effectiveness analysis.

Cost parameters can sometimes be difficult to assess, especially when individuals are lost to follow-up over time or not followed long enough for researchers to observe the event of interest, usually death. When this happens, an individual is said to have been censored. Censoring may produce biased inputs for cost and cost-effectiveness analyses because the costs that are observed will often be an underrepresentation of true health care use. Therefore, it is important for researchers to use methods that try to recover some of the information that is lost when data is censored.

It can still be possible to estimate mean lifetime, or mean interval-specific, costs when censoring is present if we make parametric assumptions about the cost data. Early efforts to produce unbiased cost inputs treated censored costs and uncensored costs as the equivalent of censored time and uncensored time. These studies applied the Kaplan–Meier method to obtain a survival function for costs, and then computed the area under the curve. But in 1997, Lin and colleagues showed that this method would still produce biased costs because it assumed that individuals who were censored and were not censored accrued costs at the same rate, when in fact they do not. Since then, novel methods for gathering unbiased costs have emerged and warrant a closer look.
The purpose of this paper is to compare how four different costing methods perform when data is censored at varying degrees, and how their estimates impact cost-effectiveness results.

4.2 Methods

I simulated health care cost data for two cohorts of individuals receiving two different treatments (A vs. B) and under different censoring conditions (10 – 50% missingness). Next, I computed mean costs for the two cohorts using four cost estimation techniques, which include two novel cost estimators: 1) average sample; 2) complete case; 3) Bang and Tsiatis; and 4) Zhao and Tian. Finally, I used the cost estimates from the four estimation techniques to conduct a cost-effectiveness analysis.

4.2.1 Simulation

Simulating Survival Times and Follow-up

Each treatment cohort was populated with 500 individuals. The cohort receiving treatment A had a mean survival time of 60 months. The cohort receiving treatment B had a mean survival time of 36 months. Each individual incurred costs for health care use and treatment, until they were lost to follow-up (censored) or experienced the event of interest (death). Although all individuals were simulated until death, the follow-up time for the cost and cost-effectiveness analysis included only the first two years (24 months). This timeline was selected to reflect cost-effectiveness analyses that often accompany randomized control trials, which have shorter follow-up times.
Cost Simulation

The cost data of individuals receiving treatment A were simulated to reflect an expensive intervention that provides some mortality benefit and the cost data of individuals receiving treatment B were simulated to reflect a less expensive intervention with less benefit. Individuals’ costs were simulated with four inputs: 1) initial cost; 2) fixed monthly cost; 3) random monthly cost; and 4) terminal cost. Costs were simulated with four inputs so that I could manipulate costs to reflect a real-world scenario in which there is an increase in health care utilization at the beginning of treatment, which plateaus over time with some variability and typically peaks again at the end of life. I modified the \texttt{ccostr} package in RStudio to simulate costs.\textsuperscript{68} The \texttt{ccostr} package in RStudio was designed to compare cost estimators with a limited prespecified dataset and did not allow for cost inputs to vary.

For individuals receiving treatment A, initial costs were sampled from a uniform distribution with a minimum of $1,000 and a maximum of $1,500. Fixed monthly costs were sampled from a uniform distribution with a minimum of $10,000 and a maximum of $20,000. Random monthly costs were sampled from a uniform distribution with a minimum of $0 and a maximum of $2,500. Terminal costs were sampled from a uniform distribution with a minimum of $10,000 and a maximum of $20,000. Costs from all four inputs were summed to calculate the total cost per individual.

For individuals receiving treatment B, initial costs were sampled from a uniform distribution with a minimum of $1,000 and a maximum of $1,500. Fixed monthly costs were sampled from a uniform distribution with a minimum of $2,500 and a maximum of $5,000. Random monthly costs were sampled from a uniform distribution with a
minimum of $0 and a maximum of $1,000. Terminal costs were sampled from a uniform distribution with a minimum of $25,000 and a maximum of $30,000. Again, costs from all four inputs were summed to calculate the total cost per individual.

Simulating Censoring

I modified the ccostr package to create five different censored datasets that increased by 10% per dataset so that the first dataset only had 10% of costs being censored and the fifth dataset had 50% of costs censored. Originally, the ccostr package did not allow for specific inputs for the percent censored and only included two options for censoring: light and heavy. Censoring was done randomly and was not meant to correlate with disease severity or treatment effect.

4.2.2 Cost Estimators

The following cost estimation techniques were used to adjust for censoring: 1) average sample; 2) complete case; 3) Bang and Tsiatis; and 4) Zhao and Tian. In the next section, I provide examples of how to compute costs under each technique. A detailed list of the variables needed for cost calculations can be found in Table 4.1.
Table 4.1 Notation and Explanation

<table>
<thead>
<tr>
<th>Notation</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$i$</td>
<td>Individual $i$; if $i = 1$ then we are referring to the first individual, furthermore if $i = 2$ then we are referring to the second individual and so forth</td>
</tr>
<tr>
<td>$t$</td>
<td>Time $t$; if $t = 1$ then we might mean at time = 1 month or 1 year, it will depend on the context</td>
</tr>
<tr>
<td>$n$</td>
<td>Total number of individuals</td>
</tr>
<tr>
<td>$T_i$</td>
<td>Survival time for individual $i$</td>
</tr>
<tr>
<td>$C_i$</td>
<td>Censored time for individual $i$</td>
</tr>
<tr>
<td>$X_i$</td>
<td>$X_i = \min(T_i, C_i)$, also referred to as the follow-up time for individual $i$</td>
</tr>
<tr>
<td>$M_i(t)$</td>
<td>Cumulative cost for patient $i$ at time $t$</td>
</tr>
<tr>
<td>$\Delta_i$</td>
<td>$I(T_i \leq C_i)$, or think of this as: Death indicator (or event of interest indicator), if $\Delta_i = 1$ then we observed the individual until death, if $\Delta_i = 0$ then we did not observe the individual until death (they were censored before we could observe their death)</td>
</tr>
<tr>
<td>$\tilde{R}(X_i)$</td>
<td>$K(t) = \Pr(C_i &gt; t)$, also referred to as the Kaplan-Meier estimate for censoring, see Bang and Tsiatis Estimator section</td>
</tr>
</tbody>
</table>
Average Sample Estimator

The average sample estimator ignores censoring completely, but it was important to include the average sample estimator in my analysis to demonstrate how ignoring censoring can produce biased costs. When using the average sample estimator, take the average of the total costs among all individuals at the end of follow-up:

\[
\text{AS}_{\text{Total Cost}} = \frac{1}{n} \sum_{i=1}^{n} M_i
\]

(1)

Using equation 1, if there are five individuals and their cumulative cost at the end of follow-up are: 10, 20, 10, 5, and 5 dollars, for individuals 1, 2, 3, 4, and 5, then \(n = 5\), and \(M_1 = 5, M_2 = 20, M_3 = 10, M_4 = 5,\) and \(M_5 = 5\). Next, sum the cumulative cost and divide the sum by 5 to get the average sample cost estimate of $10.

Complete Case Estimator

The complete case estimator deals with censoring by omitting all observations that are censored from the cost calculation:

\[
\text{CC}_{\text{Total Cost}} = \frac{1}{n^*} \sum_{i=1}^{n^*} \Delta_i M_i
\]

(2)

where \(n^*\) is the number of uncensored individuals. Using the example described previously and equation 2, let’s say that individuals 1 and 3 are censored. Then the complete case cost estimate would be computed in the following way:

\[
\frac{1}{3} [1(20) + 1(5) + 1(5)] = 10.
\]
Bang and Tsiatis Estimator\textsuperscript{69}

The Bang and Tsiatis estimator adjusts for censored data by reweighting uncensored costs by the probability that they will be censored:

$$BT_{Total\ Cost} = \frac{1}{n} \sum_{i=1}^{n} \frac{\Delta_i M_i}{\tilde{R}(T_i)}$$

where $\tilde{R}(T_i)$, or the Kaplan-Meier estimate for censoring is calculated the same way we would a Kaplan-Meier survival statistic; however, the event of interest would be censoring, not survival. I demonstrate how to calculate the Kaplan-Meier estimate for censoring in Table 4.2.
Table 4.2 Computing the Kaplan-Meier Estimate for Censoring

<table>
<thead>
<tr>
<th>t</th>
<th>Individual OP (QR#STPRU)</th>
<th>Pr(censored)</th>
<th>Pr(censored)</th>
<th>$\hat{R}(T_t)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>C A A D A</td>
<td>$\frac{1}{5}$</td>
<td>$\frac{4}{5}$</td>
<td>$1 \times \frac{4}{5} = \frac{4}{5}$</td>
</tr>
<tr>
<td>2</td>
<td>A A D</td>
<td>$\frac{0}{3}$</td>
<td>$\frac{3}{3}$</td>
<td>$\frac{4}{5} \times \frac{3}{3} = \frac{4}{5}$</td>
</tr>
<tr>
<td>3</td>
<td>A C</td>
<td>$\frac{1}{2}$</td>
<td>$\frac{1}{2}$</td>
<td>$\frac{4}{5} \times \frac{3}{2} \times \frac{1}{2} = \frac{2}{5}$</td>
</tr>
<tr>
<td>4</td>
<td>D</td>
<td>$\frac{0}{1}$</td>
<td>$\frac{1}{1}$</td>
<td>$\frac{4}{5} \times \frac{3}{2} \times \frac{1}{2} \times \frac{1}{1} = \frac{2}{5}$</td>
</tr>
</tbody>
</table>

D = died; A = alive; C = censored
Briefly, at \( t = 1 \), individual 1 is censored and individuals 2 – 5 are not. The probability of being censored at \( t = 1 \) is \( \frac{1}{5} \) and the probability of not being censored is \( \frac{4}{5} \).

The Kaplan-Meier estimate that an individual could be censored at \( C_i > T_i \) would be computed by taking the product of the probability of not having been censored previously (no one had been censored prior to \( t = 1 \)) and the probability of not being censored at \( t = 1 \): \( 1 \times \frac{4}{5} = \frac{4}{5} \). These steps are repeated at all follow-up times.

Using the Kaplan-Meier estimates for censoring from Table 4.2, and equation 3, and the costs for individuals 1 – 5 described previously, the Bang and Tsiatis cost estimate would be computed in the following way:

\[
\frac{1}{5} \left[ \frac{0+5}{5} + \frac{1+20}{5} + \frac{0+10}{5} + \frac{1+5}{5} + \frac{1+5}{5} \right] = \$12.50.
\]

The reweighting of the Bang and Tsiatis estimator makes it so that individuals who are not censored can represent a total of \( \frac{1}{R(T_i)} \) observations. Like the complete case estimator, the Bang and Tsiatis estimator still ignores costs belonging to censored individuals.

Zhao and Tian Estimator

The Zhao and Tian estimator adds to the Bang and Tsiatis estimator by incorporating the cumulative costs of censored individuals, rather than ignore them:

\[
ZT_{Total\ Cost} = \frac{1}{n} \sum_{i=1}^{n} \frac{\Delta_i M_i}{R(T_i)} + \frac{1}{n} \sum_{i=1}^{n} \frac{(1 - \Delta_i)(M_i(C_i) - \overline{M}(C_i))}{\overline{R}(C_i)} \tag{4}
\]

where \( \overline{M}(C_i) \) is the average of the cumulative costs at time \( C_i \) among individuals who are still alive at time \( C_i \).
In equation 4, notice that the first part of the Zhao and Tian estimator is the Bang and Tsiatis estimator. The second part of the Zhao and Tian estimator incorporates the cumulative costs from censored individuals and reweights them by differencing out the average cumulative costs from all other individuals still alive and dividing by the Kaplan-Meier estimate for censoring. Using the same example as before, if we assumed that the average cumulative cost of those still alive at time $C_1 = 5$ and $C_3 = 9$ dollars, then the Zhao and Tian cost estimate would be computed in the following way:

$$
\frac{1}{5} \left[ \frac{1 \times 20}{5} + \frac{1 \times 5}{4} + \frac{1 \times 5}{4} \right] + \frac{1}{5} \left[ \frac{(1 - 0)(5 - 5)}{4} + \frac{(1 - 0)(10 - 9)}{2} \right] = \$13
$$

4.2.3 Survival Analysis

Life expectancy was used to as the effectiveness measure of treatments A and B. I used the survival package in RStudio to compute and graph the Kaplan-Meier survival curves, and the restricted mean life expectancy (area under the Kaplan-Meier curve) at two years of follow-up.\textsuperscript{72}

4.2.4 Cost-Effectiveness Analysis

The true average of the simulated treatment cost was $331,724 and $94,734 for treatment A and B, respectively. At two years of follow-up, the true simulated life expectancy was 19.9 months and 18.2 months for treatment A and B, respectively. The difference in the average cost between treatment A and B was $236,990 and the difference in effectiveness, at the end of two years was 1.7. Using the true cost and effect measures described above, the true ICER of treatment A was $139,406.
I calculated the ICER, described above, by taking the difference of costs and dividing it by the difference in effectiveness between treatment A and treatment B for each simulated dataset. For each simulated dataset I calculated a percent deviance from the true ICER by taking the absolute value of the difference between the true ICER and the simulated ICER, divided by the true ICER.

4.3 Results

Figures 4.1 and 4.2 contain the average sample, complete case, Bang and Tsiatis, and Zhao and Tian cost estimates of treatments A and B at varying degrees of censoring. The vertical line represents the true cost from each cohort, which was $331,724 and $94,734 for treatments A and B, respectively. In Figure 4.3, I show the survival curves associated with treatments A and B at varying degrees of censoring. In Table 4.3, I present the cost and cost-effectiveness results under each cost estimation technique. Table 4.3 also contains information on how biased cost-effectiveness ICERs were.
Figure 4.1 Cumulative 2-Year Cost Distribution of Treatment A and B
4.3.1 Cost Estimators

Average Sample

The average sample estimator consistently underestimated the total cost of treatments A and B at all levels of censoring. Compared to the other estimators, the average sample estimator produced the most biased costs. Though all estimators underestimated the true costs of each treatment, the average sample estimator underestimated costs at higher levels than the other estimators. From 10 – 50% censoring the average sample’s 95% confidence interval (CI) of treatment A’s and B’s costs did not overlap with the true cost.

Complete Case Estimators

The complete case estimator underestimated costs at higher levels than the Bang and Tsiatis and Zhao and Tian. From 10 – 50% censoring the complete case’s 95% confidence interval (CI) of treatment A’s costs did not contain the true cost. The complete case’s 95% CI only contained the true cost of treatment B at 10% censoring.

Bang and Tsiatis Estimator

Cost estimation was slightly improved with the Bang and Tsiatis estimator. Though all costs were underestimated, the Bang and Tsiatis estimates were closer to the true mean cost of both treatments A and B when compared to the average sample and complete case estimates. I found that the Bang and Tsiatis cost estimates contained the true cost of treatment at 10% censoring. The Bang and Tsiatis cost estimates were unable to recover the true cost at higher levels of censoring (20 – 50%).
Zhao and Tian Estimator

The Zhao and Tian cost estimates were the least biased among all the cost estimators. Like the Bang and Tsiatis estimator, I found that the Zhao and Tian cost estimates contained the true cost of treatment at 10% censoring and that it was unable recover the true cost at higher levels of censoring (20 – 50%).
Figure 4.2 Cost Estimation Techniques’ Performance (Treatment A)

Cost Estimators:
- AS = Average Sample
- CC = Complete Case
- BT = Bang and Tsiatis
- ZT = Zhao and Tian

Vertical line is true cost ($331,724)
Figure 4.3 Cost Estimation Techniques’ Performance (Treatment B)

[Graph showing cost estimation techniques' performance with varying censorship rates (10%, 20%, 30%, 40%, 50%) and cost estimators (AS, CC, BT, ZT).]

Cost Estimators:
- AS = Average Sample
- CC = Complete Case
- BT = Bang and Tsiatis
- ZT = Zhao and Tian

Vertical line is true cost ($94,734)
4.3.3 Survival Analysis

The true follow-up time for treatment A and B was 386 and 262 months, and the true life expectancy at two years of follow-up were 19.9 and 18.3 months, respectively. The follow-up time of the Kaplan-Meier survivor curves varied slightly from 10 – 50% censoring. At 10% censoring the maximum follow-up time for treatment A and B was 282 and 192 months, respectively. Median survival time at 10% censoring was 47 and 27 months, respectively for treatments A and B. At 50% censoring, the maximum follow-up time for treatment A and B was 109 and 90 months, and median survival time was 48 and 26, respectively. The mean two year life expectancy remained consistent as censoring varied from 10 – 50%. At two year follow-up, mean life-expectancy reflected the true life expectancy, which was 19.9 and 18.3, for treatment A and B.
Figure 4.4 Survival Analysis

[Graphs showing survival analysis with different levels of censoring: 0%, 10%, 20%, 30%, 40%, 50% censoring.]
4.3.4 Cost-Effectiveness Analysis

Average Sample Estimator

Using the average sample estimator to estimate the costs of treatments A and B resulted in the most biased ICERs among all the cost estimators. As censoring increased, the average sample ICER estimates continued to grow more biased. From 10 – 50% censoring the average sample ICER deviated from the true ICER from 4 – 36%. Though it should be noted that the ICER still showed that treatment A was not cost-effective even at 40% of censoring, which is reflective of the true ICER.

Complete Case Estimator

The complete case estimator, like the average sample estimator, also produced biased ICER estimates as censoring increased. The complete case ICERs deviated from the true ICER by 4 – 34% under varying degrees of censoring. Still, even though cost were underestimated when compared to their true values, the cost-effectiveness analysis showed that treatment A was not cost-effective, when costs were adjusted for with the complete case estimator and at censoring levels below 50%.

Bang and Tsiatis Estimator

Using the cost inputs from the Bang and Tsiatis estimator, when censoring was between 10 – 50%, the simulated ICERs deviated from the true ICER between 4 – 31%. Like with the previously mentioned cost estimators, the cost-effectiveness analysis showed that treatment A was not cost-effective, when costs were adjusted for with the Bang and Tsiatis estimator at censoring levels below 50%.
Zhao and Tian Estimator

In the cost-effectiveness analysis, the Zhao and Tian estimator produced the least biased ICERs. The ICER point estimate, using the Zhao and Tian inputs, still identified treatment A as not cost-effective when censoring was between 10 – 40%. However, at 50% censoring the ICER point estimate for treatment A was below the $100,000 WTP threshold, making it cost-effective, when in fact it we simulated the data to show that it was not. At 50% censoring, even the Zhao and Tian estimator was not able to recover enough information to eliminate bias from our cost-effectiveness results.
Table 4.3 2-Year Cost-Effectiveness Results

<table>
<thead>
<tr>
<th></th>
<th>Treatment A Costs</th>
<th>Treatment B Costs</th>
<th>ICER</th>
<th>% Deviant</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Truth</td>
<td>Average</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$331,724</td>
<td>$99,734</td>
<td>$139,406</td>
<td></td>
</tr>
<tr>
<td>10% Censoring</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average Sample</td>
<td>319,712</td>
<td>91,944</td>
<td>133,981</td>
<td>3.9</td>
</tr>
<tr>
<td>Complete Case</td>
<td>320,547</td>
<td>92,647</td>
<td>134,059</td>
<td>3.8</td>
</tr>
<tr>
<td>Bang and Tsiatis</td>
<td>321,440</td>
<td>92,685</td>
<td>134,562</td>
<td>3.5</td>
</tr>
<tr>
<td>Zhao and Tian</td>
<td>321,815</td>
<td>92,816</td>
<td>134,705</td>
<td>3.4</td>
</tr>
<tr>
<td>20% Censoring</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average Sample</td>
<td>303,907</td>
<td>87,278</td>
<td>127,429</td>
<td>8.6</td>
</tr>
<tr>
<td>Complete Case</td>
<td>307,383</td>
<td>88,578</td>
<td>128,709</td>
<td>7.7</td>
</tr>
<tr>
<td>Bang and Tsiatis</td>
<td>309,024</td>
<td>88,859</td>
<td>129,509</td>
<td>7.1</td>
</tr>
<tr>
<td>Zhao and Tian</td>
<td>310,261</td>
<td>89,135</td>
<td>130,074</td>
<td>6.7</td>
</tr>
<tr>
<td>30% Censoring</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average Sample</td>
<td>282,268</td>
<td>82,312</td>
<td>117,621</td>
<td>15.6</td>
</tr>
<tr>
<td>Complete Case</td>
<td>286,890</td>
<td>84,283</td>
<td>119,181</td>
<td>14.5</td>
</tr>
<tr>
<td>Bang and Tsiatis</td>
<td>289,786</td>
<td>84,880</td>
<td>120,533</td>
<td>13.5</td>
</tr>
<tr>
<td>Zhao and Tian</td>
<td>291,720</td>
<td>85,287</td>
<td>121,431</td>
<td>12.9</td>
</tr>
<tr>
<td>40% Censoring</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average Sample</td>
<td>254,899</td>
<td>76,949</td>
<td>104,676</td>
<td>24.9</td>
</tr>
<tr>
<td>Complete Case</td>
<td>262,300</td>
<td>79,192</td>
<td>107,711</td>
<td>22.7</td>
</tr>
<tr>
<td>Bang and Tsiatis</td>
<td>266,383</td>
<td>79,966</td>
<td>109,657</td>
<td>21.3</td>
</tr>
<tr>
<td>Zhao and Tian</td>
<td>269,261</td>
<td>80,487</td>
<td>111,044</td>
<td>20.3</td>
</tr>
<tr>
<td>50% Censoring</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average Sample</td>
<td>222,584</td>
<td>69,620</td>
<td>89,979</td>
<td>35.5</td>
</tr>
<tr>
<td>Complete Case</td>
<td>229,725</td>
<td>72,676</td>
<td>92,382</td>
<td>33.7</td>
</tr>
<tr>
<td>Bang and Tsiatis</td>
<td>236,810</td>
<td>74,083</td>
<td>95,722</td>
<td>31.3</td>
</tr>
<tr>
<td>Zhao and Tian</td>
<td>239,382</td>
<td>74,643</td>
<td>96,905</td>
<td>30.5</td>
</tr>
</tbody>
</table>
4.4 Discussion

I compared four methods that correct for censored cost data, they included: 1) average sample; 2) complete case; 3) Bang and Tsiatis; and 4) Zhao and Tian. I found that all of the estimators underestimated the true costs of treatment. Despite this, some cost estimators were better than others. Of note, the Zhao and Tian estimator was able to reweight costs in a way that produced point estimates that were less biased than the average sample and complete case estimators. In cost-effectiveness analyses, I found that cost inputs gathered using the Zhao and Tian estimator did not produce overly biased ICERs, when censoring was below 50%.

None of the cost estimation techniques were able to recover all of the information that was lost from censored observations. This highlights the need to still include cost inputs in sensitivity analyses. A sensitivity analysis would at least provide information on the values of when treatment costs are so high that they are no longer cost-effective, so that researchers will be less likely to claim cost-effectiveness when it may not be fully warranted.

I want to acknowledge that previous work has already been done in the field of cost estimators. Three key papers have already explored the use of cost estimators on costing or cost-effectiveness studies; however, my paper differs from previous literature in the follow ways:

• Techniques for estimating health care costs with censored data: an overview for the health services researcher\(^{73}\)
The authors simulated then compared costs using the: 1) full-sample estimator; 2) uncensored case estimator; and 3) Bang and Tsiatis’ inverse probability weighted estimator, under 7, 18, 21, and 53% censoring. Furthermore, they compared the previously mentioned estimators and two additional estimators, the: 4) Bang and Tsiatis’ partitioned cost estimator; and 5) Lin’s 1997 estimator on a real dataset of patients with heart failure. The difference between my paper and Wijeysundera and colleagues’ paper is that I include a new cost estimator (the Zhao and Tian estimator). The Zhao and Tian estimator has been shown to be more accurate that the previously mentioned estimators when cost history is available. Furthermore, I include examples of how the cost estimators impact a cost-effectiveness analysis.

• On estimating medical cost and incremental cost-effectiveness ratios with censored data

This paper was written by Zhao and Tian, the developers of the Zhao and Tian estimator. The paper was written to illustrate how the Zhao and Tian estimator can be used to estimate cost parameters which are then used in a cost-effectiveness analysis. The difference between my paper and Zhao and Tian’s paper is that I include varying degrees of censoring to demonstrate how censoring impacts the cost estimators and their application in a cost-effectiveness analysis. Furthermore, Zhao and Tian did not compare their cost estimator to other cost estimators that exist. What my paper adds is that I compare the Zhao and Tian estimator to the Bang and Tsiatis estimator and two additional estimators, the average sample and complete case estimators.
• Estimation of mean health care costs and incremental cost-effectiveness ratios with possibly censored data\textsuperscript{74}

The authors describe how to use the hcost program, which was developed for STATA users, to implement the Bang and Tsiatis and Zhao and Tian estimators on a simulated dataset. Following cost estimation, the authors describe how to perform a cost-effectiveness analysis with the hcost program. The difference between my paper and Chen and colleagues’ paper is that I include varying degrees of censoring to demonstrate how censoring impacts a cost-effectiveness analysis. In addition, I include two additional estimators, as previously stated.

\textit{4.4.1 Limitations}

One limitation of this paper is that we assumed censoring was completely random. This is sometimes not the case because we know that some patient characteristics may be correlated with follow-up times. To my knowledge, no cost estimator is able to adjust for correlated follow-up times. This type of issue can resolved with study design, such as including a proactive follow-up strategy to decrease censoring for certain individuals, or advanced statistical methods that adjust for informed censoring.\textsuperscript{75-77}

A second limitation is in the setup of the cost data. I assumed that costs were low at the beginning, steady over time, and high at end of life. In situations when costs do not increase at end of life, the cost estimation techniques may not underestimate the true cost of treatment as much as was seen in my study. Future studies should consider varying cost accumulation patterns.
Furthermore, in practice cost estimates may come from multiple sources – unlike what I have shown here. For this reason, researchers may want to be extra careful as to how those costs are applied. For example, my simulations did not test for what would happen to the ICER if treatment A’s costs were adjusted for with the average sample estimator and treatment B’s costs were adjusted for with the Zhao and Tian method. Because this relationship is not well known, every effort should be made to source costs from the same study, and costs should be adjusted for with the same costing techniques to prevent further bias.

Finally, additional cost estimators exist which were not compared in this paper. Some examples include the Lin estimators and the replace from the right estimator. In 2007 Zhao and colleagues showed that the Lin estimator and the Zhao and Tian estimator can yield the same cost estimates under certain circumstances. Furthermore, in 2011, Zhao and colleagues proved that the Bang and Tsiatis and the replace from the right estimators yield the same cost estimates. Researcher should be aware that other estimators exist and that sometimes they will produce equivalent results. Researchers may wish to use multiple estimators to compare results.

4.4.2 Conclusion

Despite these limitations, this paper builds upon previous knowledge. Specifically, my findings support the use of novel cost estimation techniques. The Zhao and Tian estimator performed exceptionally well compared to less complex costing methods. Furthermore, programming packages in STATA and R exist and accompanying literature that demonstrates how to use them are now widely available. Researchers should use cost adjustors to correct for censored data in all future studies.
Chapter 5 Conclusion

This dissertation utilized decision science techniques to inform women’s health policy and practice regarding the increasing epidemic of opioid use disorder, specifically among pregnant women. In the first chapter, I identified a cost-effective screen-and-treat strategy. In chapter three, I used a value of information analysis to support future research priorities. In chapter four, I demonstrated the importance of using cost estimators to analyze censored cost data. My results are one of few that can be used to inform treatment guidelines for pregnant women with opioid use disorder, a historically understudied population, and the methods I used may serve as a practical guide for future decision scientists.

In chapter two, I used an innovative approach combining both maternal and child outcomes into one decision model to identify a cost-effective screen-and-treat strategy for pregnant women with opioid use disorder. I found that screening for opioid use disorder at the first and second prenatal visit with the 4Ps Plus and treating those who screen positive with methadone maintenance treatment was cost-effective; the incremental cost-effectiveness ratio was $28,800 per quality-adjusted life year. I found that treating maternal opioid use disorder decreased maternal morbidity, stillbirths, preterm births, and infants born with NAS by 90, 37, 27, and 45 percent, respectively, compared to no screening and no treatment. Given these results, physicians should consider using the 4Ps Plus to screen for opioid use disorder. Furthermore, decision makers should prioritize efforts to increase access to methadone to improve women’s health.

In chapter three, my value of information analysis showed that at least one additional observational study that specifically enroll pregnant women was warranted to
measure the utility of treatment on methadone and buprenorphine. I estimated that an observational study of this caliper would have a fixed cost $50,000 per person and an additional cost of $500.33 per individual enrolled. I found that the optimal sample size for the proposed observational study should be no more than 140 persons, after weighing the expected value of sample information with the cost of research. I tested two other research studies: 1) a randomized control trial designed to measure time on treatment and maternal and infant outcomes; and 2) an observational study focused on evaluating screening performance. The value of information analysis showed that these studies were not warranted, given the low value they would provide and the high cost of research.

Decision making is informed by cost parameters which represent trade-offs between different treatments; therefore, ensuring that cost estimates encapsulate all treatment related costs, including follow-up costs, is important. In chapter four, I found that the Zhao and Tian estimator for censored costs produced the least biased cost and cost-effectiveness results when compared to the average sample, complete case, and Bang and Tsiatis estimators. At 10 – 50% censoring, using cost estimates from the Zhao and Tian estimator produced the least biased ICERs, which only deviated from the true ICER between 4 – 31%. In the context of evaluating screen-and-treat strategies for opioid use disorder in pregnancy, cost adjustments made to censored data have the potential to prevent under-valuing strategies which incur costs upfront but provide downstream improvements in health. The long-term benefits of treatment for opioid use disorder include reduced risk of death and disease, and improvement in mental health, all of which would be unobserved if individuals were lost to follow-up.80
Decision science is a powerful tool used to inform decision making across a variety of topics. In this dissertation, I apply decision science techniques to the problem of opioid use disorder in pregnant women and make cost-effective recommendations on a screening tool and treatment option. Furthermore, I use a value of information analysis to identify areas of missing research that could change screen-and-treat guidelines. Finally, I test specific cost estimators to illustrate how new techniques produce less biased results and should be considered a research standard. Insights gained from this dissertation may be useful for researchers who wish to implement decision science methods in their own work.
References

1. Klein E. In the UK’s health system, rationing isn’t a dirty word. *Vox* 2020.
45. Smith K LRWoCAaOdp.
63. VOI Gaussian Approximation.  