

**Incorporating Comparative Effectiveness Research Results into the U.S. Medicare
Program**

Samuel T. Savitz

Table of Contents

I.	Introduction	Pg. 3
II.	Literature Review	Pg. 6
III.	Methodology	Pg. 10
IV.	Analysis of CER Articles for Heart Stents	Pg. 14
V.	Analysis of CER Articles for Knee-Arthroplasty	Pg. 18
VI.	Analysis of CER Articles on Prostate Cancer	Pg. 21
VII.	Analysis of National Medicare Data on Prostate Cancer	Pg. 26
VIII.	Analysis of Intermountain Prostate Cancer Data	Pg. 35
IX.	Case Study on NICE	Pg. 47
X.	Case Study on the Oregon Medicaid System	Pg. 52
XI.	Health Polling Results	Pg. 56
XII.	Conclusion and Policy Recommendations	Pg. 58

Section I: Introduction

Comparative Effectiveness Research (CER) is a method of evaluation that compares the effectiveness of two or more medical treatments by assessing comparative health outcomes. Historically, there has been a lack of implementation of the findings from CER. Typically, the approval process for a new treatment judges it on the criterion of whether it works efficaciously with little to no consideration of the relative effectiveness with existing treatments [1]. This process for approving new treatments fails to measure how treatments of a similar type perform relative to one another and does not account for the costs of alternative treatments.

More recently, there has been a growing movement towards advancing CER in the United States with research investment coming from the federal government, private insurance plans, and the pharmaceutical industry. Recent legislation has also promoted CER. The Medicare Modernization Act of 2003 provided funding for the Agency for Healthcare Research and Quality (AHRQ) to study comparative effectiveness. Also, the American Recovery and Reinvestment Act of 2009 (ARRA) apportioned \$1.1 billion for CER, with \$300 million going to AHRQ, \$400 million going to The National Institutes of Health, and \$400 million going to the Department of Health and Human Services [2]. Further, the Institute of Medicine report listed the 1546 requests for different studies of CER in the U.S., which suggests that the demand for this type of information is considerable [3].

This federally funded research prohibits inclusion of cost-effectiveness in the analysis [4]. The Center for Medicare and Medicaid Studies has explicitly stated that they will not use cost considerations or the results from CER to make coverage decisions, which is likely due to public opposition. The Center for Medicare and Medicaid Service (CMS) claims to cover any treatment that is “reasonable and necessary for the diagnosis or treatment of illness or

injury”. Though the meaning of this phrase is ambiguous, it does imply that costs will be a criterion [5].

Researchers performing the CER studies sometimes incorporate the cost-effectiveness of the treatments as part of the analysis. Most other developed countries in the world use CER that includes cost-effectiveness in making national coverage decisions [6]. When cost-effectiveness is included as a criterion, the result is usually expressed as a ratio of a monetary unit over the health benefits. The most common measure for health benefits in CER is quality adjusted life years (QALYs) [5]. The analysis will detail both CER with and without the cost-effectiveness criterion. When discussing CER with the cost-effectiveness criterion, I will make a notification.

Without the cost-effectiveness criterion, CER would determine which treatment is most beneficial in terms of health outcomes (like pain, side-effects, and successful treatment of the condition) for the population being considered no matter how expensive the treatment or how marginal the benefit relative to alternative treatments that are cheaper.

Another issue is whether national coverage decisions and payment rates would incorporate the CER findings. The inclusion of research results is important for the impact that the research will actually have. If results are not incorporated into coverage decisions and evidence-based practice guidelines, then reliance would be shifted onto physicians to implement the findings voluntarily. But if physicians have ingrained treatment practices or they are not up to date on the current literature, physicians would probably not implement these results into their practice.

The recent Patient Protection and Affordable Care Act provides additional funding for CER and features an Independent Payment Advisory Board (IPAB) that will address CER in the Medicare system. [7-9] The Board consists of fifteen members that make proposals for ways that

Medicare can reduce the expenditure in the Medicare system if growth exceeds certain measures. The legislation prohibits the Board from making proposals that ration care; increase revenues; change benefits, eligibility or Medicare beneficiary cost sharing (including Medicare Part A and B premiums); or would result in a change in the beneficiary premium percentage or Part D low-income subsidies [10]. Though, the Bowles-Simpson deficit reduction plan proposes a far stronger role for the IPAB and so future legislation may remove some of these restrictions [11].

Proponents of CER assert that better research and application of the research through an organization like the IPAB would have several beneficial effects. One such impact would be a determination of which treatments provide the most health improvements in particular situations [2, 12]. If the dissemination of these results were successful, then physicians would employ more effective treatments for their patients. Health quality would improve as a result of these more effective treatments.

Supporters of these policies also claim that CER would reduce Medicare expenditures by reducing wasteful spending. These reductions would accrue for patients, employers, and third-party payers (insurance companies and the government) because they would not be paying less for health care that qualifies as wasteful spending [13]. However, this claim is not self-evident, since CER would not necessarily lead to decreases in health spending in all cases. It is conceivable that CER results could increase the nation's Medicare expenditure by promoting coverage for and use of expensive treatments that were determined in the absence of cost-effectiveness considerations.

My research will focus on how the IPAB can promote more incorporation of CER findings into practice within the Medicare system. To do so, I will examine the current state of the CER literature, the public acceptance of CER, and the need for CER in clinical practice.

Section II. Literature Review

The literature suggests that the effects of CER depend greatly on how physicians and health systems implement the results of these studies. The U.S. healthcare system has many barriers in place that deter incorporation of CER, though there are notable opportunities as well.

Barriers to implementation of CER:

There is controversy as to how much of the care in the United States is not cost-effective. The proportion of care that qualifies as cost-effective depends largely on the calculations. The most subjective aspect of this calculation is the specific value of the threshold under consideration. A high cost-effectiveness threshold would mean that fewer treatments would qualify as cost-ineffective, while a low cost-effectiveness threshold would have the opposite effect. Whether there is a significant reduction in health expenditure depends in large part on how the level for this threshold is established. Further, new treatments have come at a reasonable value (<\$50,000/QALY is the lowest commonly cited threshold in cost-effectiveness studies), so it is already the minority of treatments that a cost-effectiveness criterion would target [14]. Even the majority of care in a stereotypically high cost area of care like cancer treatment comes at a reasonable value [15]. Several studies in the United States suggest that if a cost-effectiveness threshold were to exist, it would definitely exceed \$50,000/QALY and could be over \$100,000/QALY using estimates from contingent valuation and revealed preferences [14, 16]. Thus, the commonly used \$50,000/QALY ratio may greatly overestimate the amount of care that is cost-ineffective.

The National Institute for Clinical Excellence (NICE) is an independent body that makes coverage recommendations for the National Health Service in England and Wales. Their analyses consider costs as well as the benefits of particular treatments [5]. In the case of NICE,

cost-effectiveness is not the only criterion under consideration in their research. NICE considers values like fairness and equity and may justify approval of a treatment that exceeds the threshold for cost-effectiveness as a result [17]. I will explain more thoroughly NICE relates to the IPAB in Section IX. Determining the proportion of care that is not cost-effective does not take these important values into consideration and it may be useful to continue certain treatments that are above the threshold.

Even with CER including cost-effectiveness as a criterion, the costs for the Medicare program would not necessarily decrease in all cases. The cost-effectiveness criterion does not relate directly to aggregate health spending [5, 16]. While CER with cost-effectiveness does identify treatments that are not cost-effective and are being over utilized, it also indicates many treatments that are cost-effective and underutilized [16, 18-20]. Therefore, the IPAB would not save costs in all cases like those when it encourages more treatment that meets the cost-effectiveness standard.

If CER were to be implemented without the cost-effectiveness criterion, then the CER findings would support health improvement regardless of how much they cost. An example of a hypothetical scenario that highlights a potential problem with this approach appears in the table below with a comparison of two alternative treatments:

	Cost	QALYs
Treatment A	\$10,000	1.0
Treatment B	\$100,000	1.1

According to CER without consideration of cost-effectiveness, Treatment B is the best option because it results in more QALYs than Treatment A. Thus, CER without the cost-effectiveness

criterion could increase costs by encouraging the more expensive treatment. Further, most new treatments represent at least a minor improvement in care, hence few treatments would be rejected under this criterion [1]. Given the context of rapidly increasing health costs, it would likely be infeasible for the IPAB to accept such treatments that could result in significantly higher expenditure for Medicare.

If the Medicare payment rate for physicians does not reflect the CER studies, then it is less likely that physicians would incorporate the research into their practices. Physicians often react to payment incentives and so they will not necessarily change how they treat patients when research indicates that one treatment is more effective than another [21]. However, making coverage decisions and changing the payment rates may not be politically feasible options for IMAB.

Opportunities for implementation of CER:

The primary area where CER would apply is care that provides little or no benefit. This means that the treatment does not result in improved quality of life, life expectancy, or only slight improvements in either measure. The care that provides no benefit or even harm is often called wasteful spending and some estimates put the percentage of care that is wasteful as high as thirty percent of all health care spending, which amounts to over \$400 billion in wasteful per year [22]. Further, the cost-effectiveness for treatment varies greatly between different patient groups [23]. For example, the average cost per year of life gained among the elderly is several times as great as it is for the non-elderly. This suggests that the average treatment for the elderly is providing much less benefit per dollar spent and that the IPAB could target more care in the Medicare system [14]. Additionally, many studies indicate that higher expenditures do not have a positive correlation with better health outcomes [4, 20, 23-26]. One possible explanation for this

pattern is that there is excess care of little or no benefit or care that may even be harmful. While, this correlation alone does not demonstrate any such causal relationships, the utilization of low-benefit care likely plays a role.

If such wasteful spending does exist in the ways listed above, then for some of the care, CER with cost-effectiveness would be able to demonstrate that the cost of the treatment is very high relative to the marginal benefits as compared with alternatives. The IPAB might then reject or limit the treatment under these circumstances if this does not conflict with the political limitations for this group. Even if the IPAB is unable to recommend eliminating coverage of wasteful care on cost-effectiveness grounds, it will be able to do so if the CER findings show that the clinical effectiveness of other treatments surpasses this care [10].

IPAB could also use its role to challenge biases and incentives physicians have not to incorporate the results of CER studies. Physicians have a tendency to choose the more complex and expensive treatments for several reasons. First, there is a bias among physicians towards using newer or more complex treatments without any evidence that they are more effective [7, 14, 25]. Further, more complex and expensive procedures will often lead to greater payment for the physicians, so they have incentives to perform them [25, 27]. Finally, there is a phenomenon called technology creep for when physicians gradually apply technology that is cost-effective in a sub-group to the whole group without justification [28]. New treatments, especially treatments incorporating new health technology, usually cost more and have been a major driver of the increase in Medicare expenditure [13, 14, 23, 24, 28]. Thus, a greater emphasis on CER with or without cost-effectiveness for coverage decisions would limit new and more expensive treatments that are not significantly better than the existing treatments [13, 29, 30].

In summary the literature suggests uncertainty over how the IPAB could incorporate CER into the U.S. healthcare system to improve health outcomes and potentially decrease health expenditures. Wasteful spending (care that exceeds a reasonable cost-effectiveness criterion) seems prevalent in the US system, and CER could promote care that is more cost-effective. However, it is difficult to identify which treatments or areas of care are contributing to this problem and, hence, it is uncertain how well CER can address this kind of spending. Applying CER may also increase the costs in certain areas of the healthcare system, which may not be feasible given the current concerns over costs in the Medicare system. Finally, the ability to address costs may depend on whether IPAB can actually use the cost-effectiveness criterion.

Section III. Methodology

My primary research question asks how the IPAB can successfully incorporate CER into clinical practice decisions within the Medicare system. I will evaluate this question by looking at the data for three specific medical conditions that will serve as case studies. Specifically, I will examine: 1) the use of stents in the treatment coronary artery disease; 2) knee-arthroplasty; and 3) prostate cancer treatment. I chose these three conditions because they each have extensive CER literature and there are reasonable alternative treatments. Further, the conditions are different in meaningful ways that will make my findings more generalizable. One such way is that intervention is far less discretionary for a patient in need of treatment for a coronary artery blockage, which is life-threatening in an immediate sense [31]. Also, the preliminary literature review suggested that the evidence of clinical benefit for treatments differs significantly between prostate cancer and the other conditions with prostate cancer having much less clarity and many more options [31-33]. Finally, mortality is not a concern with failing to provide knee-arthroscopy

whereas it may occur with failure to treat the other conditions; knee function is purely a quality of life matter [34].

Research Questions and Plans:

The table below lists the questions that follow from my main research question. Specifically, the questions in the table address how extensive the CER literature for these treatments and what impact they already have on clinical practice and government policy. Determining the answers to these questions will help answer what kind of role the IPAB can achieve in promoting CER for these conditions. I will partially answer the first four questions (1-4 in fig. 1) by examining the CER literature on treatments for these three conditions. I will conduct a systematic literature review of these three conditions using the online Duke Library database search tool. I will also research best practice guidelines to answer the fifth question. I can find these best practice guidelines from medical organizations like American Academy of Family Physicians. The best practice guidelines will help confirm the accuracy of my findings from my own literature search on these conditions and supplement missing information.

Figure 1: Research Questions

	Heart Stents	Knee Arthroplasty	Prostate Cancer Treatment
1. What are the findings from the CER literature?			
2. How well do Medicare reimbursements accord with the CER?			

3. Is there underutilization or overutilization of care			
4. Are there variables besides patient preferences and clinical need that influence care?			
5. How well do best practice guidelines match the CER?			

I will also answer the third and fourth question through analysis of Medicare data. I will perform separate regressions on patient data (sources explained below) to answer the questions relating to the provision of care for prostate cancer. One reason for limiting the analysis to just prostate cancer is that the treatments for this condition have the most uncertainty. Further, similar access to data sources did not exist for the other two conditions. Lastly, limitations on time necessitated reducing the scope of the research.

Data Sources:

The first data source is the peer reviewed literature. Secondly, state-level data from the Dartmouth Atlas of Health Care and the Healthcare Cost & Utilization Project (HCUP) detail hospital-level analysis of Medicare claims data. The variables include treatment type (just for prostate cancer), region, hospital characteristics, reimbursement rates, and information about Medicare beneficiaries that utilize that hospital. I will analyze this data using frequency tables and regressions. Significant differences of the distribution of treatments over hospital characteristics or region could suggest a lack of clear evidence or a lack of implementation and a need for more research.

The final data source is a HIPAA-compliant analysis of de-identified, patient-level data from Intermountain Healthcare on the recurrence of prostate cancer. The purpose of this analysis is to determine how well alternative treatments for prostate cancer work as judged by the recurrence rate. All other variables being equal, a higher recurrence rate would suggest that one of the alternative treatments has worse outcomes. This latter focus is aligned with the interests and needs of the Oncology Clinical Program; work will be supervised by Mr. Tom Belnap, Senior Outcomes Analyst in the Institute for Health Care Delivery Research.

The model for this approach becomes iteratively more specific with each step. The first step details the overview of the literature on the subject. The second step builds upon the first by showing state-level data on treatment for prostate cancer. Finally the third step drills down to the patient level and functions as a comparative-effectiveness study.

Case Studies:

I will supplement this research on procedures in the United States with two short case studies of how the National Health Service (NHS) in England and Wales incorporates CER into coverage decisions and how the Oregon Medicaid experiment utilized research in the early 1990's. Specifically, I will examine how the NHS uses CER from a policy perspective instead of doing an in-depth study of the effects that such a policy has on their health care system. This information will indicate how the United States government could feasibly implement CER in the Medicare program. To do so I will need to examine the exact details how the NHS makes coverage decisions and the exact ways they employ CER. I can find this information through literature on the subject. Likewise, for the Oregon Medicaid example, I will need to determine how the Oregon government used research and how the public responded. This case will show how the public reacted to the use of cost-effectiveness to determine coverage decisions. This

example will also suggest what the IPAB could feasibly do given the constraints within the American system.

In both these cases external validity is the main concern. I will explain why the lessons from the cases studies actually apply to what the IPAB can do for the whole Medicare system.

Section IV. Analysis of CER Articles for Heart Stents

The literature review focused on the CER for drug eluting stents (DES) against bare metal stents (BMS). Using the search terms “coronary artery blockage comparative effectiveness” and limiting my search to English language, I identified 30 articles for review. Review of the articles yielded five studies. An additional seventeen articles came from the bibliographies of these five papers. I summarize my findings in the next sections.

Findings from CER Literature:

The general consensus in the literature held that DES had better outcomes than BMS. The findings for heart stents was informative concerning the advantages and disadvantages of each option. However, uncertainties about thrombosis and off-label use persist and they need further study for a more thorough evaluation of the alternatives.

The advantages of DES include lower restenosis rates and potentially decreased mortality. Restenosis means that the vessel that physicians operated on has become restricted again. All studies I examined showed that the restenosis rate was significantly lower in DES than for BMS [35-49]. Thus, this conclusion has a high degree of certainty.

Other potential advantages for DES include lower mortality rates and decreased incidence of cardiac events. However, only a minority of studies found statistically significant differences [45, 48-50], so this finding may reflect the truth, but it currently lacks sufficient

support. Most of these studies did not look beyond one year of treatment, so these studies would likely not determine long-term trends in mortality and cardiac events.

Though, one possible advantage for BMS would be their impact on late thrombosis. Thrombosis means the formation of a blood clot in a blood vessel. One of the studies found a statistically significant increase in the rate of thrombosis for DES [51]. This study is a meta-analysis so it has a high degree of credibility since it draws from the findings of multiple articles. One explanation for why other studies do not find a statistically significant increase in thrombosis is their short time range: the meta-analysis only found this statistically significant result with the inclusion of late thrombosis (after one year).

The economic evaluations in this review do not present a clear advantage for the cost-effectiveness of DES compared to BMS. While the clinical advantage belongs to DES, the extent of this advantage remains unclear. These cost-effectiveness studies did not include the potential impact of thrombosis on the quality of life. If a significant difference exists between the two treatments for rates of thrombosis, then these studies would overestimate the clinical advantage of DES. The studies report the following findings: \$27,450/QALY, >\$50,000/QALY, and \$58,721/QALY (last value is in Canadian dollars). These ratios are the incremental cost-effectiveness ratios (ICERs) for DES. However, certain sub-groups performed better with DES like those patients over 65, small vessels, and other high-risk groups and the resulting ICER was higher. Also, two of the three studies took place outside the U.S. and so they may be less applicable. The studies also found that the cost-effectiveness ratios were sensitive to the difference in costs between the two kinds of stents. These cost-effectiveness estimates suggest that DES, but concerns remain over the applicability of the studies and the non-inclusion of thrombosis [43, 45, 47].

The last major consideration concerns improper use of stents. About 60% of stents go towards off-label use, so it is important to understand how the different treatments apply towards these individuals and not just towards on-label use, which most studies look at. Off-label use means that the patients do not have the right indications for receiving the stents. These patients have a higher average number of co-morbidities and tend to be sicker. But DES still performs better for these patients in an observational study, with lower rates of restenosis [52]. However, I did not find a randomized experiment looking at off-label stent use, so this conclusion lacks strength because of selection concerns. Also, compliance with treatment is an issue because patients with DES need to take anti-platelet therapy for several years after implantation. Patients that do not comply have increased rates of restenosis and thrombosis [38, 39, 53, 54]. Though, the information on non-compliers may be valuable to include since there will also be non-compliers in practice. If patients had higher (or lower) rates of compliance than in these studies, the outcome differences between DES and BMS would alter in favor of DES (BMS).

Areas for Further Study:

These studies lack information on certain outcomes that may affect preferences among treatments and makes the comparative effectiveness of these treatments less accurate. Specifically, the connection between DES and increased rates of late thrombosis should be examined more carefully. A higher rate of late thrombosis may affect patients' decisions. Also, more investigation of the differences between DES and BMS on mortality and cardiac events over a longer time period would help inform treatment. Without information on these complications, the relative advantages of these alternatives are not fully known. Lastly, more cost-effectiveness studies in the U.S. would help make a better estimation of the cost-effectiveness ratio for DES to see if health gains over BMS come at a reasonable price. These

studies should also incorporate the findings from these other studies on thrombosis, mortality, and cardiac events as they develop to obtain a more accurate picture of the health benefits.

Best Practice Guidelines for Heart Stents:

The best practice guidelines for heart stents confirm the lack of absolute clarity found in the CER review. The American College of Cardiology (ACC) and the American Heart Association (AHA) state that DES and BMS during angioplasty are both viable interventions in most situations. These organizations stated that there is currently no evidence suggesting differential effects on major health outcomes except for DES decreasing the rates of restenosis. However, ACC and AHA recommend that in cases where the patient may not fulfill their anti-platelet therapy, BMS should be used [55]. More CER studies could potentially allow the ACC and AHA to have more specific conclusions.

NICE differs slightly in its advice, though it reaches the same general conclusions as the article review and the ACC and AHA guidelines. It addresses cost-effectiveness, which the ACC and AHA recommendations do not fully consider. They conclude that DES does lower the rate of restenosis. However, the treatment which has the better ICER differs depending on the price differential between the two types of stents and patient characteristics like diabetes co-morbidity or having small vessels. Thus, the cost-effectiveness vary depending on these factors [56]. Presumably, NICE would benefit from having better CER findings to use when making their recommendations.

Both of these best practice guidelines substantiate my findings on the relative benefits of DES in comparison to BMS. They both find DES better than BMS in terms of restenosis. Also, in both cases they do not mention off-label stents, nor do they come to a conclusion regarding

late thrombosis, which confirms the lack of information that also appears in the literature [55, 56].

Section V. Analysis of CER Articles for Knee-Arthroplasty

I focus on the comparison between unicompartmental knee-arthroplasty and total knee-arthroplasty in reviewing the literature. Using search terms “osteoarthritis comparative effectiveness knee-replacement” and “rheumatoid arthritis comparative effectiveness knee-replacement” and limiting my search to English language, I identified 64 articles for review. Review of abstracts yielded five studies, excluding three as irrelevant on closer inspection. Three additional articles came from reviewing the bibliographies of the other papers. Below I summarize my findings.

Findings from CER Literature:

The CER literature for knee-arthroplasty contains less breadth of information than the literature for heart stents. Despite this limitation, I did find clear findings relating to the advantages and disadvantages of the treatment options.

Nearly all studies find that total knee arthroplasty (TKA) and unicompartmental knee arthroplasty (UKA) are effective treatments for addressing loss of knee function from arthritis [57-60]. The treatments differ in that the surgical procedure for TKA applies to a much larger proportion of the knee than does the procedure for UKA. Though these studies indicate that the two treatment options are effective, they do not clearly demonstrate that one treatment works comparatively better.

The evidence suggests that the two treatments are approximately equivalent since there are few statistically significant differences. But this only applies to cases where both treatments

could potentially repair the knee joint. Sometimes the arthritis has spread to so much of the knee that UKA would not work well. Several studies suggest that UKA leads to better range of motion, fewer complications, and faster recovery than TKA in situations where both would be appropriate [58-60]. Though, revision rates tend to be higher for UKA as compared to TKA [58, 60]. Given that each treatment has negative results associated with it, it is difficult to make a determination on which one works better overall. One cost-effectiveness evaluation in my review had UKA with a slight advantage over TKA in terms of QALYs [58] and another study had TKA with a slight disadvantage [60].

Given the uncertainty regarding the QALY measurement for the cost-effectiveness ratios, the two main economic analyses provide different results. The first study has UKA dominating TKA [58]. Dominating means that one treatment has better outcomes and saves money compared to another. The other study has an ICER of \$65,245/QALY for TKA as compared to UKA. But this ICER is sensitive to costs of the two treatments and relevant co-morbidities [60].

The incidence of UKA is much lower than TKA despite evidence that it may be more or just as effective in many cases. A study on all knee-replacements in the U.S. from 1998-2005 showed that only 8% of knee-replacements were UKAs and the rest were TKAs. While UKA is not appropriate in all cases, it is applicable in many more than 8% of the cases. This low percentage suggests that some bias exists against UKA in favor of TKA. It is possible that the evidence suggesting that UKA yields similar results in comparison to TKA is relatively recent and so the shift may not take place for some time. It is important to note that the growth of UKA during this time period amounted to 32.5%, which is much higher than the 9% growth for TKA [61]. Therefore, a slow trend exists towards a greater proportion of UKAs during this time period. But this fact still does not fully explain why there is not greater use of UKA. Perhaps

physicians have become accustomed to performing TKA and do not want to change their practice patterns. It is also possible that the higher reimbursement rates for TKA relative to UKA may incentivize the former treatment [58].

Areas for Further Research:

The most important area of research relates to the uncertainty regarding the advantages UKA and TKA. A more detailed examination of the quality of life criteria would help determine the relative benefits for each procedure. Specifically, researchers should develop better estimates for range of motion improvements, complications, recovery time, and revision to make comparing the two treatments easier.

The difference in cost-effectiveness between UKA and TKA illustrates the importance of accurate measures. These cost-effectiveness studies have relatively similar values for the quality of life improvements for UKA and TKA patients, but since they have different treatments having a slight advantage in health outcomes, they lead to very different conclusions. If UKA did not have the advantage in QALYs in the one case, then it would not dominate TKA. This is an example of how slight changes in the QALYs can lead to significant implications for the cost-effectiveness of treatments.

Finally, there should be studies on the reasons why the percentage of UKA receivers is so low compared to TKA receivers. Understanding why this occurs will help policymakers that want to encourage more UKA procedures. Doing so could reduce Medicare expenditure without sacrificing health.

Best Practice Guidelines for Knee-Arthroplasty:

The guidelines for knee-arthroplasty offer clarity for treatment options, though they also note gaps in the research. The Osteoarthritis Research Society International (OARSI) provides a

detailed best practice guideline for this condition. The guideline advises that people suffering from osteoarthritis in their knees should first try aerobic exercise and muscle strengthening. They also recommend that overweight patients lose weight and take corticosteroids. If patients do not find sufficient symptom relief from these options in combination with pharmacological interventions, OARSI counsels people to seek joint-replacement surgery. The researchers point out that TKA and UKA are both effective treatments. Though, these guidelines also note the absence of conclusive findings that determine which treatment is preferable [62].

NICE provides similar recommendations on treatment. They advise that clinicians should initially provide non-surgical treatments like exercise. Clinicians should utilize surgery for those patients who do not respond to these treatments. NICE also finds that UKA and TKA are both suitable treatments without delivering conclusions on which is more appropriate in different situations [63].

Both of these best-practice guidelines confirm my findings from the review of CER studies for knee-arthroplasty. In each case, they find that both UKA and TKA are acceptable, but that further CER is necessary to better understand the relative advantages and disadvantages.

Section VI. Analysis of CER Articles on Prostate Cancer

The literature on prostate cancer treatment suggests a great deal of uncertainty for the comparative benefits of each treatment. Using search terms “prostate cancer comparative effectiveness” and limiting my search to English language, I identified 30 articles for review. Review of abstracts yielded 14 studies, excluding two as irrelevant on closer inspection. Below I summarize my findings.

Findings from CER Literature:

The analysis of the literature suggests uncertainties for the relative costs and benefits of the main treatment. These alternatives consist of watchful waiting, radical prostatectomy, robot-assisted prostatectomy, external beam radiation, laparoscopic prostatectomy, retropubic prostatectomy, and androgen deprivation therapy. Further research is necessary to make more useful comparisons between these options.

A great deal of variation exists between the relative proportions of these treatments in different settings. One would expect them not to vary so greatly if it is just clinical need and patient/physician preferences determine treatment. For instance, all types of prostatectomies ranged between 11-82% across U.S. hospitals. Further, watchful waiting varied from between 0 and 28% across hospitals. Such extreme variation cannot be due to preferences alone, but rather must involve some bias within these hospitals or regions towards certain treatments. Further, the uncertainty between treatment options is so great that clinical effectiveness does not serve as a useful guide for treatment [64]. Interestingly, the newest procedures tended to have the highest growth rates, suggesting that adoption had a lot to do with the procedure that seemed the most advanced [65].

Certain advantages appear to exist between the various types of prostatectomy. For instance, laparoscopic prostatectomies have an association with shorter stays and bladder/urethral obstruction rates than open prostatectomies [66, 67]. Another study found higher rate of future cancer treatment for laparoscopic prostatectomies [68]. Also, minimally invasive techniques are sometimes related to higher rates of erectile dysfunction and genitourinary complications [69]. While advertisements make robot assisted prostatectomy appear to be the favored treatment, the only significant benefits include lower blood loss and transfusion rates

during surgery [70, 71]. Despite these relationships, stronger evidence is necessary before determining that one type of prostatectomy has better clinical outcomes than another.

There are also many studies suggesting advantages and disadvantages of different types of prostatectomy and radiation therapy. Prostatectomy has higher average rates of urinary dysfunction, though radiation therapy has higher rates of bowel disorders [72-74]. Sexual outcomes varied, but there was only one study from 1995 that suggested higher rates of sexual dysfunction side-effects for prostatectomy [72-74]. And rates of recurrence do not appear to have a relationship with treatment type [75]. So there does not seem to be a clear preference for one treatment type over another (i.e., radiation vs. prostatectomy) since both have advantages and disadvantages.

Similar to the previous treatments, watchful waiting has advantages and disadvantages relative to other treatment types. Notably, patients do not have to undergo treatment and face the risk of treatment related side-effects [76]. In this comparison, it is reasonable for patients to prefer no treatment to avoid these side-effects. Though, one study did find a lower mortality rate for men randomized to radical prostatectomy [77], In the U.S., prostate specific antigen (PSA) scores detect many patients as having prostate cancer vs. detection happening when symptoms started. In fact, some of these patients would have never developed symptoms before death. These results may not apply to this group that has less severe prostate cancer among members that test positive by their PSA levels. It is difficult to obtain good results on watchful waiting patients because the long-term effects happen over a period of many years [76].

The cost-effectiveness evaluations face the major limitation of the uncertainty in the clinical benefits for these different treatments. One study looks at just the costs between the different treatments because the benefits cannot be ascertained. This cost comparison has radical

prostatectomy as being \$487 cheaper than laparoscopic prostatectomy and \$1726 cheaper than robotic assisted prostatectomy. The difference with robotic assisted prostatectomy stayed large even when the initial cost of the robot was omitted (\$1155). The authors concluded that laparoscopic may be a competing alternative to radical prostatectomy with similar costs and possible benefits. However, they stated that with only reduced hospital days and blood loss these benefits are not enough to offset the much higher costs. Unless the initial cost of the machine or maintenance and use costs decrease, robotic assisted prostatectomy would likely not have a favorable incremental cost-effectiveness ratio [70].

Areas for Further Research:

The uncertainties related to prostate cancer treatment suggest many areas where future comparative effectiveness research would be useful. For instance, more long-term research on how watchful waiting compares to more aggressive treatment would help clarify the advantages of each alternative. Though, it is difficult to obtain good findings when the long-term outcomes ten years into treatment are so important. Also, more CER is needed on how the newer forms of prostatectomy compare to the traditional procedures. An interesting topic from a policy perspective is how robotic assisted prostatectomy compares to other, much cheaper and traditional procedures. The concern is that robotic assisted prostatectomy potentially offers little or no clinical benefits, yet physicians choose to use this option because they get a higher reimbursement. Such behavior occurs in relation to IMRT, a type of radiation treatment for prostate cancer patients. Medicare pays up to \$40,400 for each patient, which has led urologists to purchase their own equipment and self-refer their patients. Once they have the machine they then face a huge financial incentive to refer their own patients to get this expensive treatment [78]. Other explanations include physicians trying to get their money's worth out of the machine

investment or favoritism towards newer advances. These studies must also take into account the learning curve that physicians face when using new treatments like robotic assisted prostatectomy, which may explain the lack of observed benefits [71]. More research in these areas would help create a better understanding of the relative benefits of each treatment type.

Best Practice Guidelines for Prostate Cancer:

The best practice guidelines for localized prostate cancer confirm the uncertainty I found in the literature. The American Urological Association (AUA) provides a comprehensive best practice guideline for this condition. The guideline states that brachytherapy, external beam radiotherapy, watchful waiting, and radical prostatectomy are all effective treatments that lead to similar results. The analysis of the data does not show any significant differences in patient outcomes, though for high risk patients they suggest active therapy may be preferable because of the increased likelihood of mortality for watchful waiting. Except for this qualification, the AUA recommends focusing on patients' preferences when making the decisions without specifying which treatment is more appropriate in different cases [79].

NICE offers similar recommendations on the treatment of localized prostate cancer. For low-risk patients, NICE advises watchful waiting. This does differ from the other guideline and conforms with the tendency for NICE to consider costs when making recommendations. NICE states that the other treatments are acceptable options, just not the preferred ones. For moderate- to high-risk patients, NICE advocates use of radical prostatectomy or radiotherapy. Likewise, NICE notes that patients' preferences regarding potential side effects should have a role in the decision-making process [80].

Both guidelines note the lack of definitive evidence for comparing treatment options. The AUA points out that "study outcomes data do not provide clear-cut evidence for the superiority

of any one treatment” for low risk patients [79]. Further, they note a lack of randomized experiments establishing the comparative- and cost-effectiveness of some of the treatments. While NICE does provide recommendations, they still lack clarity on the advantages for different forms of active treatment [80]. These guidelines support the finding from the literature review that more CER would be useful for addressing these weaknesses in the available evidence. If nothing else, the CER could support patient preferences by determining what the likelihoods of particular side-effects are.

Section VII. Analysis of National Medicare Data for Prostate Cancer

The analysis of the Medicare data for prostate cancer treatment helps to confirm my findings on the extent of variation for prostate cancer treatment and suggests possible explanations for the variation. An understanding of the type of factors that cause undue variation in treatment practices would aid the IPAB in its efforts to promote CER. For instance, the IPAB could target hospitals whose treatment practice departs from the CER literature the most. This information would also help the IPAB understand what barriers exist that prevent successful incorporation of CER. As I noted earlier, I chose to only examine prostate cancer because it is the condition that has the most variation.

To analyze the data, I fitted an Ordinary Least Squares (OLS) model using both forward and backward stepwise selection. Forward selection determines which variables to include in the model based on whether adding them contributes significantly to the model. Alternatively, backward selection starts with a model with all variables and takes out those that do not contribute significantly to the model. I had difficulty interpreting the regressions since the number of states was only 31. This low number of observations meant that inclusion of more

than two variables in the model would cause the standard errors to inflate greatly. In turn, this would make it so that neither the model nor the coefficients for the variables would be statistically significant. To still get interpretable results, I used several different methods for the forward and backward stepwise selection.

The data entries come from state level data. The reason the number of observations is only 31 is that only 33 states had data in the HCUP database and two states did not have information on the racial breakdown of those patients receiving prostatectomy treatment. HCUP is an all-payer, claims-based dataset maintained by the Agency for Healthcare Research and Quality with voluntary state reporting. The states that participated and had the information on race include: AZ, AK, CA, CO, FL, HI, IA, KS, KY, ME, MD, MA, MI, MS, NE, NV, NH, NJ, NY, NC, OK, OR, RI, SC, TN, TX, UT, VT, WA, WI, and WY.

Explanation of Variables:

The variables consist of measures for prostatectomy, characteristics of the population receiving prostatectomy, and information about the hospitals and Medicare systems in each state. Information on each variable in the analysis follows below:

1. The number of prostatectomies per 1000 male Medicare beneficiaries in 2008. This variable is the outcome variable in the models. The analysis centers on looking for which independent variables have a significant relationship with this variable. The reason that I do not look at other treatments is that the Medicare coding did not provide data on watchful waiting (Medicare does not reimburse for this) and the number of patients were too low to examine the alternatives.
2. The ratio of prostatectomy in nonteaching hospitals to prostatectomy in teaching hospitals in 2008. This independent variable could have a relationship with the outcome

because the different types of hospitals have distinct practice patterns or they incorporate research to a different extent.

3. The ratio of prostatectomy in non-metropolitan hospitals to prostatectomy in metropolitan hospitals in 2008. This variable could have an association with the outcome variable because non-metropolitan hospitals may have different financial incentives than metropolitan areas that affect treatment patterns and adoption of best practices.
4. The proportion of hospitals that have large number of beds in 2008 data (definition of having a large number of beds varies depending on location, urban status and teaching status; for specific cutoffs see Appendix A). Large number of beds may have a relationship with the quantity of prostatectomy because a larger number of beds would suggest that the hospitals in states that more frequently perform prostatectomy tend to have a high capacity. This higher capacity could influence the type and number of patients that receive care.
5. The number of inpatient days for the average Medicare beneficiary during the last two years prior to death from 2001 to 2005. This variable may have an association with the outcome variable because the number of inpatient days is a proxy for treatment intensity. More intense treatment in general would suggest more aggressive treatment of prostate cancer.
6. The number of surgical procedures/1,000 Medicare beneficiaries in 2005. The number of surgical procedures may relate to the rate of prostatectomy because states that have higher surgery rates overall would likely have higher rates of prostatectomy.

7. The dollar amount of the mean charge for prostatectomy that Medicare insured. This variable may have an association with the prostatectomy rate because higher charges could indicate a monetary incentive for physicians and the hospital to do prostatectomy.
8. The dollar amount for the average reimbursement across a state for Medicare beneficiaries during the last two years of life between 2001 and 2005. This variable may relate to the outcome because higher reimbursement levels suggest more treatment. Also, higher reimbursement rates could induce more treatment because they work as a financial incentive.
9. The U.S. Census Region for each state. There are four U.S. Census Regions.
10. The ratio of the proportion of blacks that received prostatectomy to the proportion of blacks in the Medicare system in that state in 2008. This variable shows whether the proportion of blacks receiving prostatectomy is unusual for that state adjusting for the different proportions of blacks in states. The reason for including this variable is that blacks have a higher incidence of prostate cancer than other races. Without this variable, a state that had a particularly high proportion of black residents may appear to have a higher overall rate of prostatectomy, even though it is primarily due to this overrepresentation of a high risk group.

Model One; Forward Selection Considering Original, Squared, and Square Root:

This model comes from forward selection considering all variables in their original form, in their squared form, and in their square root form. The variables that the selection method produced include the square root of the number of inpatient days and the ratio of teaching to non-teaching hospitals. The r-squared value is 0.2902 and the p-value for the F-test is 0.008 and is significant at the 0.05 significance level.

Prostatectomies/thousand = $-.0008315 * (\text{square root of the average number of inpatient days}) +$
 $-.000103 * (\text{ratio of non-teaching to teaching hospitals}) + 0.005531$

States with higher average numbers of inpatient days tend to have lower ratio of prostatectomies per thousand Medicare beneficiaries. Also, the lower the ratio of non-teaching to teaching hospitals in the state (that is, a lower proportion of non-teaching hospital) correlates with a lower ratio of prostatectomies per thousand Medicare beneficiaries.

Model Two; Backward Selection Considering Original, Squared and Square Root:

This model comes from a backward selection considering all variables in their original form, in their squared form, and in their square root form. The variables included in the model are the square root of the ratio of procedures done in metropolitan vs. non-metropolitan hospitals and the average Medicare reimbursements. The r-squared value is 0.2215 and the p-value for the F-test is 0.03, which is significant at the 0.05 significance level.

Prostatectomies/thousand = $-.0006073 * (\text{square root of ratio of non-metropolitan to metropolitan hospitals}) +$
 $-4.61e-08 * (\text{Medicare reimbursement in last two years of life}) + 0.004097$

States with higher with larger ratios of non-metropolitan to metropolitan hospitals or a smaller average Medicare reimbursement for the last two years of life tend to have a lower rate of prostatectomies.

Model Three; OLS for Original Variables Individually:

I performed regressions of all variables in their original form against the outcome variable. I regressed each variable one at a time to determine significance. Only the variable of inpatient days was significant with a p-value of 0.028. The r-squared is 0.1471 and the p-value for the F-test is 0.0276, which is significant at a 0.05 level.

Prostatectomies/thousand = $-.0000762 * (\text{the average number of inpatient days}) + 0.003226$

States having higher average inpatient days was associated with a lower rate of prostatectomies.

Model Four; Backward and Forward Selection for Interactions:

This model comes from both the forward and backward selection for the interaction terms (both provided the same variables). The interaction terms were not part of the previous stepwise selections because there was too much collinearity to run the selection with these variables included. The first interaction term was the number of inpatient days times the Medicare reimbursement. The second interaction term was the ratio of metropolitan to non-metropolitan hospitals times the proportion of large hospitals. The r-squared value is 0.2986 and the p-value for the F-test is 0.007, which is significant at the 0.05 level. However, the plot showing (the ratio of metropolitan to non-metropolitan hospitals)*(the proportion of large hospitals to non-large hospitals) on the x-axis and prostatectomies/thousand on the y-axis suggests that there may be leverage points (see Appendix B). After removing these points from the model, (the ratio of metropolitan to non-metropolitan hospitals)*(the proportion of large hospitals to non-large hospitals) is no longer significant at a 0.05 significance level. Thus, the trend is not statistically significant without the inclusion of these two high leverage states. Further analysis of these states is necessary to determine whether there may have been measurement error.

$$\text{Prostatectomies/thousand} = -1.37e-09 * \text{int3} + -.0014356 * \text{int28} + .003148$$

Int3= (the average number of inpatient days)*(the Medicare reimbursement for the last two years of life)

Int28= (the ratio of metropolitan to non-metropolitan hospitals)*(the ratio of large hospitals to non-large hospitals)

States that had a larger interaction term of inpatient days times the Medicare reimbursement for the last two years of life or the ratio of metropolitan to non-metropolitan times the ratio of non-large hospitals tend to have lower rates of prostatectomy.

Model Five; Backward Selection Considering Original and Log:

This model comes from a backward selection of the natural logs of each variable and the variables in their original form. This selection method gave the log of inpatient days and the ratio of teaching to non-teaching hospitals as the significant variables. The r-squared value is 0.3027 and the F-test has a p-value of 0.0045, which is significant at the 0.05 level.

Prostatectomies/thousand = $-.001831(\text{the log of the average number of inpatient days}) +$
 $-.0001017*(\text{the ratio of non-teaching to teaching hospitals}) + .003148$

States that had higher number of inpatients days and larger ratios of non-teaching to teaching hospitals had a correlation with lower prostatectomy rates. The summary of significant variables in the different models appears below:

Model	1st Significant Variable	2nd Significant Variable
One	Negative square root of inpatient days	Negative ratio of non-teaching to teaching hospitals
Two	Negative square root of non-metropolitan to metropolitan hospitals	Negative Medicare reimbursement for prostatectomy
Three	Negative inpatient days	

Four	Negative Int3= (the average number of inpatient days)*(the Medicare reimbursement for prostatectomies)	Negative Int28= (the ratio of metropolitan to non-metropolitan hospitals)*(the proportion of large hospitals to non-large hospitals)
Five	Negative log of inpatient days	Negative ratio of non-teaching to teaching hospitals

Conclusions on National Medicare Data for Prostate Cancer:

Several of the variables under consideration have an apparent relationship with the rate of prostatectomies; however, the r-squares are quite low at nearly 30% explanatory power. The number of inpatient days seems to have the strongest relation. This variable or one of its different forms appears in four of the five models. Other variables that are in the model at least once include the ratio of teaching to non-teaching hospitals, the Medicare reimbursement in the last two years of life, the ratio of metropolitan to non-metropolitan, and the ratio of hospitals with a large number of beds to those with a small number of beds. While there are reasons to believe that these variables have a causal relation (see variable definitions above), the limitations of the data set make it difficult to conclude that these are actually causal relationships.

Instead, the conclusion is that these variables have a relationship with the rate of prostatectomies, which implies that there is significant variation depending on the hospital and community characteristics listed above. Further studies should look at these variables to determine if a causal relation does indeed exist. Regardless, the implication is that patient preferences and clinical need are not the sole determinants of the treatment patients receive. Also, these results suggest that CER could have an important impact if it can clarify the relative

advantages of the different treatment alternatives. These relationships with hospital type may be due in part to the fact that research is not conclusive and there is no reason to think that patient care is better or worse at hospitals that tend to have higher rates of particular treatments like prostatectomy.

One surprising result was that all the measures of utilization, like inpatient days, had a negative relationship with the rate of prostatectomies. It would make more sense if hospitals with such variables would have higher rates of prostatectomies, since this would also indicate greater utilization. A possible explanation is that hospitals with higher utilization tend to have other types of active therapy like radiation replacing prostatectomy.

Limitations:

This data set suffers from several limitations that make the conclusions weaker. For one, the regressions just provide ecological correlations, which tend to exaggerate the effects on the outcome variable. Ecological correlations refer to the use of averages over groups as individual observations. Further, because the number of observations was so low ($n=31$ states), the regression models could not have more than two variables in them without causing the variables to become non-significant. Thus, I could not create a comprehensive model. Lastly, this data set did not have information on patient outcomes. Hospitals that have more inpatient days tend to do fewer prostatectomies, but this is not necessarily a problem. The health outcomes may be just as good for patients at other hospitals with alternative treatments. But it is impossible to know without looking at how patients fare with the different interventions. Further analyses should examine data sets that include patient level data, patient health outcomes, and more observations. The next data set helps to rectify some of these limitations.

Section VIII. Analysis of Intermountain Prostate Cancer Data

To establish an even thorough understanding of treatment practice associated with prostate cancer I look at data from Intermountain healthcare. More knowledge about the results of treatments for prostate cancer would be useful for the IPAB in promoting CER. Also, this analysis is a simplification of what researchers study in CER studies. Thus, this investigation will demonstrate the relevant considerations for CER investigations [81].

Intermountain Healthcare is a fully integrated delivery system based in Utah with one facility in southern Idaho. The health system operates 24 hospitals, 130 outpatient clinics, and employs nearly 1,000 physicians with an additional 2,000+ highly affiliated physicians. The system serves nearly 60% of the 2 million Utah residents.

These records address some of the problems found in the previous analysis. Notably, the Intermountain data is on the patient-level and has information on health outcomes. With this information, I can evaluate the clinical benefit of the different treatments and I avoid the problem of using averages of large groups of people.

A de-identified, patient-level dataset of all patients diagnosed with prostate cancer at an Intermountain Healthcare facility was created. To evaluate this dataset, I used both logistic and OLS regression models. Logistic models have an indicator variable as the outcome variable and produce odds ratios for each of the independent variables. The odds ratio signifies how that variable affects the likelihood of the outcome variable taking a value of one. An odds ratio of one suggests that the variable does not impact this probability. But a value of two, for example, implies that a one unit increase in the independent variable doubles the likelihood that the outcome variable will be one. Equivalently, an odds ratio of 0.5 means that the probability has decreased by 50%. An odds ratio is statistically significant if the 95% confidence interval does not overlap with one.

Variables:

The variables in this data set encompass patient characteristics, cancer type and grade, and complications. Descriptions of each variable used in the analysis appear below:

1. **PSA recurrence:** Indicator variable for whether there is recurrence based on the Prostate-Specific Antigen test. A value of >0.1 for surgery and surgery + radiation or an increase of two points over baseline levels for radiation will yield a value of one for this indicator variable. Watchful waiting does not have a recurrence definition for this variable. This is one of the outcome variables for a logistic model because the rate of recurrence is one way of determining if the intervention was successful.
2. **Side-effects:** Indicator variable for whether there are side-effects for that patient. If the patient has any of the side-effects, they will have a value of one for the corresponding indicator variable, and this indicator variable will be one regardless of how many side-effects the person has. This is another outcome variable for a logistic model because I can also use this variable to analyze the treatments and estimate how these treatments affect quality of life.
3. **Recurrence:** Indicator variable for whether there is recurrence based on the Prostate-Specific Antigen test. A value of >0.1 for surgery and surgery + radiation or an increase of two for radiation will yield a value of one for this indicator variable. However, for this variable, I consider watchful waiting to have a 'recurrence' if the change is greater than ten points over the value at diagnosis. PSA values tend to increase with patient age, but an increase of ten generally will lead physicians to start active treatment. I am separating this from the previous recurrence variable since I

- created this variable from the data, whereas PSA recurrence was its own variable that existed before my analysis.
4. **Difference:** Continuous variable for the difference between the follow-up PSA value and the initial value at diagnosis. This is an outcome variable and I am including it because the PSA value serves as a proxy for the severity of the cancer and so changes in the PSA values will suggest either improvements or worsening of the condition. I adjusted this variable to account for the fact that the PSA flag for watchful waiting is an increase of 10 points, while recurrence for radiation is an increase of two points. So all watchful waiting differences are divided by five. Values for this variable are only available for radiation treatment and watchful waiting.
 5. **Age:** Age at diagnosis of prostate cancer. This is a control variable in the models, since age may impact the PSA value and complications.
 6. **Stage:** Cancer stage. There are different stages depending on where in the body the cancer has spread to and this measure includes four stages. This variable is also a control variable in the models since how far the cancer has spread may affect the PSA value and whether there are complications.
 7. **Diagnostic PSA value:** PSA value at admission. This controls for potentially greater differences in PSA values when the initial value is larger.
 8. **Primary Gleason:** The primary Gleason score measures the pattern of the majority of the tumor in terms of increasing severity from one to five. This is a control variable because it evaluates the severity of the condition.
 9. **Secondary Gleason:** The secondary Gleason score measures the pattern of the minority of the tumor (has to be at least five %) with the same categories as the

primary Gleason score. The reasoning is also the same for using this variable as a control.

10. Treatment: Indicator variables for the treatment category. Treatment1 is for radiation, Treatment2 is for surgery, Treatment3 is for surgery + radiation, and Treatment4 is for watchful waiting. The treatment types and corresponding number of patients appear in the table below:

Treatment	Patients
Radiation	459
Surgery	1176
Surgery + Radiation	50
Watchful Waiting	206

11. Insurance: Indicator variables for the insurance provider. I use the indicator for Medicare because I am researching the implications for Medicare.

12-25. A full set of indicator variables for common side-effects of treatment is tested.

These include:

- Radiation colitis
- Anorectal fistula
- Rectal or anal hemorrhage
- Rectal or anal disorder
- Gastrointestinal bleeding
- Cystitis NOS
- Bladder fistula

- Hemorrhage of the bladder
- Bladder disorder NOS
- Hematuria
- Radiotherapy reaction
- Stoma formation
- Surgical repair complication
- Late-effect radiation

Model one; Logistic Model Using PSA Recurrence:

The first model is a logistic model using PSA recurrence as the outcome variable. Watchful waiting is not included because the PSA recurrence measure does not apply. I also excluded radiation + surgery because the low number of patients in this category was causing absurd estimates.

The first column reports the crude odds ratio, which is a model with the only independent variable being the indicator variable for the treatment. In this case the crude odds ratio suggests that surgery has significantly higher odds ratios since the 95% confidence intervals do not overlap with one. In other words, having surgery increases the likelihood for a PSA recurrence for the estimate for the crude odds ratio. These results hold over all the sets of controls. The first set of controls is all relevant variables without regard for whether they are significant. The second set of controls only includes those that significantly contribute to the model at the 0.05 significance level. Finally, the last set of controls only includes those variables that either increase or decrease the odds ratio by 10%, which is the standard measurement for significance when using logistic models.

While I reserve the larger implications of these findings for the conclusion section, I determine that the results provide significant evidence that surgery leads to a higher probability of PSA recurrence in this sample because the odds ratios for surgery are significant in all four logistic models considered. Again, odds ratios are significant when the 95% confidence interval does not overlap with one. Also, there is significant evidence that radiation leads to a lower probability of recurrence with significant odds ratios in all four cases. It should also be noted that these are actually two separate models: each row represents the results for using the indicator variable specific to that treatment. This presentation makes it easier to understand the difference between the two treatments.

Treatment	Crude Odds Ratio	Adj. Odds Ratio	Control Variables	Adj. Odds Ratio (Stepwise sign. controls)	Stepwise sign. Controls	Odds Ratio (10% odds ratio sign. controls)	10% odds ratio sign. Controls
Radiation 16/197=0.081	0.13 (0.07-0.22)	0.09 (0.05-0.17)	Cancer Stage, Primary Gleason, Secondary Gleason, Age, Insurance	0.09 (0.05-0.18)	Gleason Secondary, Cancer Stage, Age	0.08 (0.04-0.15)	Primary Gleason, Secondary Gleason, Insurance, and Age
Surgery 187/456=0.41	7.86 (4.56-13.55)	11.09 (5.74-21.42)	Cancer Stage, Primary Gleason, Secondary Gleason, Age, Insurance	10.69 (5.59-20.41)	Gleason Secondary, Cancer Stage, Age	11.09 (5.74-21.42)	Cancer Stage, Primary Gleason, Secondary Gleason, Insurance, and Age

Model Two; Logistic Model for Side-Effects:

The second analysis is a logistic model using an indicator for whether there are any side-effects. It includes all of the treatments in the data set.

The columns and types of controls in each part are the equivalent as in the previous model, adjusted for the controls that are actually significant in this model. Radiation serves as the baseline against which I measure the odds ratios for the other treatments. In the first column the crude odds ratio is only significant for surgery, with the point estimate being 0.69 and a p-value of 0.024. None of the adjusted odds ratios are statistically significant, though surgery is close with a p-value of 0.066. For the adjusted odds ratios with the controls from the stepwise regression surgery is again statistically significant and yields a p-value of 0.006. Finally, for the adjusted odds ratios using the controls that alter the odds ratios by 10%, surgery is significant with a p-value of 0.041.

Given these results, I make the basic conclusion that there is significant evidence that surgery has an odds ratio below one, which implies that there is a lower probability for side-effects. Further, there is not statistically significant evidence that surgery + radiation or watchful waiting alters the likelihood of side-effects.

Treatment	Crude Odds Ratio	Adj. Odds Ratio	Control Variables	Adj. Odds Ratio (Stepwise sign. controls)	Stepwise sign. controls	Odds Ratio (10% odds ratio sign. controls)	10% odds ratio sign. Controls
Radiation 65/459=0.14	Baseline	Baseline	Cancer Stage, Primary Gleason, Secondary Gleason, Age, Insurance	Baseline	Cancer Stage	Baseline	Cancer Stage, Primary Gleason, Secondary Gleason, Insurance
Surgery 120/1176=0.10	0.69 (0.5-0.95)	.7 (0.48-1.02)	Cancer Stage, Primary Gleason, Secondary Gleason,	0.63 (0.45-0.87)	Cancer Stage	0.68 (0.47-0.98)	Cancer Stage, Primary Gleason, Secondary Gleason,

			Age, Insurance				Insurance
Radiation+ Surgery 10/50=0.2	1.52 (0.72- 3.18)	1.46 (0.62- 3.47)	Cancer Stage, Primary Gleason, Secondary Gleason, Age, Insurance	1.2 (0.54- 2.7)	Cancer Stage	1.34 (0.57- 3.13)	Cancer Stage, Primary Gleason, Secondary Gleason, Insurance
Watchful Waiting 31/206=0.15	1.07 (0.68- 1.71)	1.31 (0.63- 2.73)	Cancer Stage, Primary Gleason, Secondary Gleason, Age, Insurance	1.08 (0.55- 2.13)	Cancer Stage	1.31 (0.63- 2.74)	Cancer Stage, Primary Gleason, Secondary Gleason, Insurance

Third Model; OLS Regression on PSA Differences:

The third model is an OLS regression using adjusted PSA differences as the outcome variable. Again, this variable only takes values for radiation and watchful waiting.

Using a full set of controls, treatment type is a statistically significant coefficient. The regression includes a full set of controls (age, stage, diagnostic PSA, primary and secondary Gleason, diagnostic PSA value, and Medicare) and the coefficient takes a value of 12.46 (0.93-24). The p-value is 0.034, which is statistically significant at the 0.05 level. From this model, I determine that there is statistically significant evidence suggesting that watchful waiting has a relationship with increases in the adjusted PSA-difference. However, the R-squared value was only 0.11, which is quite low.

I also ran a model using just the statistically significant controls from backward stepwise selection. This model had cancer stage and diagnostic PSA value as the controls. The coefficient

on watchful waiting was 17.3 (3.6-31.01) with a statistically significant p-value of 0.014. This finding provides statistically significant evidence that watchful waiting is associated with larger differences in PSA values. Further, the R-squared value increases to 0.29, which is an improvement upon the previous model.

Fourth Model; Logistic Regression for PSA Flag:

The fourth model is a logistic model using a modified measure of recurrence that adds in watchful waiting by classifying increases of ten points for PSA levels as a ‘recurrence’. I did this because Mr. Belnap reported that this is the measure that physicians use in the Intermountain system to determine when watchful waiting patients should undergo active treatment. The treatments are then compared to each other by their impact on this variable. I excluded surgery + radiation because it was giving peculiar results due to having so few people in the sample.

The results below show that the odds ratio for surgery is statistically significant across all four models. At the same time, the odds ratio for watchful waiting is not statistically significant in any of the four models. Given these findings, I conclude that there is significant evidence that the true odds ratio for surgery differs from one and that there is not significant evidence that the odds ratio for watchful waiting differs from one.

Treatment	Crude Odds Ratio	Adj. Odds Ratio	Control Variables	Adj. Odds Ratio (Stepwise sign. controls)	Stepwise sign. controls	Odds Ratio (10% odds ratio sign. controls)	10% odds ratio sign. Controls
Radiation 16/197=0.081	Baseline	Baseline	Cancer Stage, Primary Gleason, Secondary Gleason,	Baseline	Age, Stage, Secondary Gleason	Baseline	Cancer Stage

			Age, Insurance				
Surgery 187/456=0.41	7.53 (4.36- 13)	9.76 (5.06- 18.02)	Cancer Stage, Primary Gleason, Secondary Gleason, Age, Insurance	9.47 (4.97- 18.03)	Age, Stage, Secondary Gleason	7.87 (4.38- 14.14)	Cancer Stage
Watchful Waiting 9/88=0.10	1.2 (0.51- 2.84)	1.1 (0.3- 3.98)	Cancer Stage, Primary Gleason, Secondary Gleason, Age, Insurance	1 (0.28- 3.57)	Age, Stage, Secondary Gleason	1.12 (0.35- 3.56)	Cancer Stage

Discussion of Analysis:

The primary conclusion from this data set is that watchful waiting compares well to the other treatments. There is not significant evidence that watchful waiting has a relationship with a greater likelihood of having side-effects over this time period. In this case, side-effects are a measure of quality of life, so the implication is that watchful waiting does not result in significantly worse outcomes for quality of life as measured by this proxy. Further, watchful waiting does not have a higher likelihood of requiring a change to a different treatment category over this time period. Though, watchful waiting does have a relationship with higher adjusted PSA values than radiation therapy. However, this measure may not be a meaningful one. Given this information, there does not appear to be significant evidence that watchful waiting is worse than these other treatments. However, the lack of decrease in the probability of side-effects is notable. One of the reasons someone might choose watchful waiting is because they would not occur the increased probability of side-effects from active treatment. This finding may have to do

with limitations of side-effects as a measure for quality of life, which I address in the limitations section.

Also, surgery has two statistically significant relationships. It has an association with an increased probability of recurrence and a lower likelihood of having side-effects. Given that one of these effects is a good outcome and one is a poor outcome, I cannot determine whether surgery appears to be a better or worse treatment option than the others. These factors should be part of the patient and physicians' considerations for when the patient will undergo some form of treatment for prostate cancer.

The limitations below do weaken these conclusions considerably. Therefore, I take these findings as evidence towards relationships that would require further substantiation before presenting them to guide policy or practice.

Limitations for this Analysis:

The main weakness for this analysis, that is pertinent for all models, is external validity. One issue is that Intermountain is a distinctive health system that is a national leader in terms of efficiency and quality of care. Also, there may be a particular physician culture or ability at this delivery system that separates it from many other systems of care. Intermountain is located in Utah, which may have different health behaviors that affect prostate cancer treatment adherence. Thus, some assumptions need to be made to extrapolate from what happens in a population treated at Intermountain to what happens in other health system across the United States [82].

Also, the fact that the treatments were not randomized suggests a potential problem with selection bias. I did control for several variables that may influence severity, but it is possible that there are unobservable factors that impact the relative success of the different treatments. For instance, I only had BMI figures for a minority of patients so I could not use it as a control

variable. Also, other co-morbidities like diabetes or high blood pressure would be useful controls that I lacked.

For the first model, I assume that the PSA recurrence values that Intermountain chooses accurately measure equivalent failure of the treatment. That is to say that comparing the proportion of PSA recurrence for radiation vs. surgery is a fair comparison to make. It is possible that PSA recurrence for radiation is worse in terms of health than recurrence for surgery. There may also be more concern with recurrence for one of the treatments, so the threshold is intentionally set low. In such cases, this comparison would not be legitimate. It is probably that these issues play a role, but that for the most part recurrence serves as good evaluator of treatment success. PSA recurrence is an indication that prior treatment has failed and results in a re-evaluation of the patient's condition and potentially leads to further cancer treatment. This is an important factor that influences patient management and can potentially affect outcomes. However, a PSA recurrence is an intermediate outcome and does not necessarily translate to decreased patient survival [83].

The major concern for the second model is how equivalent different side-effects are. I assume that someone with GI bleeding and someone with a bladder disorder are equivalent and that people with two side-effects suffer as much as someone with one. These assumptions are necessary to compare treatments using an indicator variable for whether someone has *any* side-effect(s) or no side-effects. It is likely that the different side-effects influence quality of life to greater or lesser degrees and that two side-effects would be worse than just one (though very few patients in the sample had two or more side-effects). Still, this measure does approximate quality of life differences since the biggest difference will be between people who have no side-effects at all and those who have at least one.

The main issue with the third model is the outcome variable, adjusted difference in PSA scores. One concern is about whether dividing the difference for watchful waiting by five makes the two variables equivalent. This makes some sense given the thresholds physicians use, but may not mean these differences are comparable across different treatments. Finally, PSA scores are not a perfect measure for severity as I mention with the first model. People may have worse prostate cancer that does not raise their PSA scores to abnormal levels [83]. However, these are in the minority of cases. Also, there may be more variables that contribute significantly to the model, but the statistical power seemed particularly low, with only one of the controls being statistically significant.

Model four has the same issues as the previous model and the first model since it is using the recurrence rate from the first model and a 'recurrence' for watchful waiting using differences in PSA value of ten or greater.

Section IX. Case Study on NICE

The next two sections are qualitative analyses of organizations that used CER or other types of research as a basis for policy decisions. These two case studies provide lessons for what the IPAB should do in its effort to promote CER and what limitations it would likely face.

The role of the National Institute for Clinical Excellence (NICE) in the National Health Service (NHS) provides additional guidance for how a similar organization could operate in the United States. The NHS is the health system in the United Kingdom, government financed, and covers all citizens. NICE is a separate organization and part of its purpose is to review new drugs, technologies, and interventions to determine whether the NHS in England and Wales

should cover them. The legislature then votes on proposed changes, though they very rarely vote against the recommendation of NICE [84].

NICE considers economic, clinical, and societal values to make its recommendations. The clinical evidence usually consists of research comparing the clinical benefits of what is under consideration against existing treatments if there are any. The economic evaluation primarily relates to the CER for conditions with cost-effectiveness [85, 86]. The last major consideration is of societal values like equality and fairness. NICE could alter their evaluation depending on the populations that stand to benefit or lose from the proposed drug, intervention, or technology [17]. For example, if a drug would primarily benefit a group of the population that is impoverished, NICE may take their socioeconomic status into consideration when formulating their decision to promote more equality.

CER with cost-effectiveness is one of the main tools that NICE uses to evaluate treatments. Cost-effectiveness incorporates the costs in the numerator and the benefits in the denominator. The costs and benefits come from an economic evaluation of health interventions. Typically, studies report the cost-effectiveness ratio in terms of the cost per quality adjusted life years (QALYs) [17, 85]. Officially NICE does not have a strict cutoff for the cost-effectiveness necessary for approving a treatment. This makes sense given that NICE considers other values besides cost-effectiveness like equality and fairness [17]. However, observation of the treatments that NICE rejects suggests that there is a general threshold that holds in most cases. This threshold ranges from 20,000-30,000 pounds per QALY (currently about \$31,000-46,000) [87].

Protest against their recommendations is a major problem for NICE. For instance, when NICE denied an expensive kidney cancer drug, Sutent, in August 2008, a political backlash resulted. This drug had a best-case, incremental cost-effectiveness ratio of \$139,800-

335,000/QALY [88]. Oncologists voiced criticisms and ordinary citizens expressed displeasure with the ruling. NICE responded by creating a new policy that allowed approval for new life-saving drugs that would be applicable for fewer than 7,000 patients per year. This change allowed the approval of Sutent, among other drugs, and thus avoided some controversy [89]. NICE officially approved Sutent on March 26th, 2009 [90]. However, this act appears inconsistent since NICE does not approve other drugs with that high of a cost-effectiveness ratio in most other situations. Also, NICE cannot elude controversy for drugs and other forms of treatment that apply to significantly larger portions of the population [89].

Another issue for NICE is criticisms of their methodology for determining recommendations. One common charge is that there are faults with the studies that NICE considers. Given that NICE uses cost-effectiveness as its main criterion, slight changes in either the costs or benefits of the treatment could change whether it falls within or over the loose threshold. Critics claim that it is, thus, somewhat arbitrary to accept or reject treatments when there is so much uncertainty about the evaluation [85]. Others point out that there are systematic biases that lead to greater rejection in some areas, like cancer treatment [89]. This means that NICE has to be especially careful to only consider studies that use appropriate methodology so there is as much certainty as possible from the findings and opponents have less material for criticism.

There are enough similarities between this case and the IPAB to justify using it as an example for how to operate. Both entities have as their purpose the limitation of treatment to promote value in their health systems using CER. Also, I am examining how the IPAB would apply to Medicare, which is a government funded health program, and the NHS is also a

government funded program. Finally, both organizations must have their respective legislative bodies approve their recommendations [9, 84].

Though it is important to note that while the broad objectives and the role of these organizations are the same, there are important differences that limit this comparison. The specifics of the Medicare and the NHS are not exactly alike. For example, culture regarding healthcare differs greatly between the countries. People in the UK appear to have a greater acceptance of rationing than in the United States, or at least rationing by cost-effectiveness [91]. Thus, NICE likely has more support for limiting care using cost-effectiveness as the main criterion than the IPAB would have. There are also differences in how the health care systems operate. Medicare applies to only elderly (at least age 65) and the permanently disabled [92], while the NHS covers all citizens [93]. Thus, the recommendations of NICE affect the entire population, while IPAB will only impact this subset. These limitations suggest that it would be unfeasible for the IPAB to simply copy NICE. They are operating in different contexts, so the role of IPAB needs to be specific to its unique context. Though, the similarities are sufficient to provide lessons that inform the IPAB about how it can perform its role.

One lesson from the NICE example is the need for accurate information. This lesson is also present in the Oregon Medicaid example presented below. Limitation of care is controversial, and critics will be looking for any faults in the research. The IPAB should be selective in the studies they utilize so that critics will have less to find fault with. This may imply that the IPAB will have to refrain from making recommendations on treatments that do not have extensive and/or high quality research. This in turn could limit the role of the IPAB, but it would have legitimacy for the treatments that did have sufficient evidence on which to make

recommendations. This will require a replicable basis for determining study quality and a threshold for evidence sufficiency that goes beyond current grade standards.

Another lesson from NICE is that there will be cases where denial of treatment conflicts with public sentiment. The case of Sutent in the UK is just one of many times where there has been an outcry from those who stand to benefit from a rejected treatment [89]. This issue is difficult to combat since stories about people suffering because they do not receive treatment are often more compelling than the argument that covering such treatments does not make sense for budgetary, health, and economic reasons. In this example, the NHS gave in to public sentiment. This is not a problem if it is a sporadic event, but in general the limitation of care cannot be effective if the NHS yields to public demand every time because it will have implications for credibility and resource allocation. Likewise, for the IPAB to have an actual effect, it must be able to make recommendations and not have them nullified simply because the public sentiment is not in accordance.

The last lesson is that engaging the public generates support for a program. NICE has been open with the information it uses to make decisions. Anyone can access the research and recommendations NICE makes on its web page [84]. However, some critics maintain that NICE does not release all of the economic information they use to make their recommendations. Citizens in England and Wales want greater transparency and would trust the organization more when they perceive their recommendations as having a secure basis [94]. NICE also has an advisory board composed of ordinary citizens that can influence recommendations [17]. When considering how the IPAB should operate, the desire for transparency and input from ordinary citizens is relevant. American citizens will likely want these same conditions and will express more support of the organization if it realizes them.

Section X. Case Study on the Oregon Medicaid System

The Oregon Medicaid experiment is a useful example of the limitation of services by a government health care program in the U.S. The lessons from this policy help inform the kind of role that IPAB can achieve. The advantage that this case study has over the NICE example is that it does a better job of addressing how Americans would respond to the IPAB limiting treatments. Though, the Oregon Medicaid system did not directly use CER to make decisions. However, this concern does not weaken the findings significantly because Oregon was still using a similar method for making its decisions.

The Oregon Medicaid experiment originated over concerns about policies resulting in the denial of care. Specifically, people worried over the unwillingness of the Oregon Medicaid system to cover a bone marrow transplant for a child in 1987. The Oregon legislature decided not to cover this treatment and they did so without fully considering cost-effectiveness literature. This case was part of the reason that the Oregon government created a more systematic approach to make decisions on which treatments to cover [95].

Further, legislators wanted to cut costs to the Medicaid program that was an increasing burden to the state. The traditional means of limiting costs in Medicaid are reducing the number of people that qualify for benefits and reducing the reimbursements to doctors. However, excluding individuals limits their access and so does the fact that fewer doctors accept Medicaid patients when the reimbursement rate decreases [95, 96]. Oregon did increase the number of Medicaid beneficiaries by 120,000 and decided to reduce expenditures by limiting the number of treatments that would be covered services [95].

The legislature tasked the Oregon Health Services Commission (HSC) to determine the treatment priorities. The HSC created a priority list of condition-treatment pairs by incorporating

cost, improvement in quality of life, and the number of years that improvement would last. An example of a condition treatment pair is prostatectomy and prostate cancer. Coverage of prostatectomy for prostate cancer does not guarantee coverage of other treatments for prostate cancer. The HSC consulted citizens in the process of determining improvements in quality of life. However, the original ranking led to illogical conclusions about how certain condition-treatment pairs compared to each other [95]. Subsequent rankings relied less on cost data and more on five year survival [97]. Once the HSC created the priority list, the legislature would determine how many of the listed treatments they would cover. The legislature had to follow the rankings when making decisions [95].

The reason the findings from the Oregon Medicaid example are applicable to the IPAB is that they deal with government-run health programs in the United States. Also, in both cases, limitation of treatments through scientific research plays a significant role. Again, these methodologies for determining which treatments to cover differed from CER in that they were comparing all treatments together and the Medicaid program serves poor Americans, who may have less political power than the elderly on Medicare. But, these similarities are sufficient to justify using the Oregon example to inform what strategies may work for the IPAB.

The first lesson is the importance of having accurate and consistent information. There was criticism against the original priority list with the illogical ranking order for some treatments. The reason for these puzzling conclusions was the inaccuracy of the information Oregon officials used to create the priority list and their particular way of making calculations. Their data approximations and methodology differed greatly from economic evaluations from the rest of the U.S. as a result. For instance, the original list did not discount future costs and the costs estimates were poor. Also, their measure of five year survival rate did not take into account

the difference between dying immediately and dying four years after treatment. One response by Oregon was to de-emphasize cost considerations as a criterion for later priority lists and instead focus on survival [98].

To avoid this problem, the IPAB would need to ensure that the CER studies they consider represent high quality research. The IPAB is in a politically precarious situation, and the use of faulty information would severely damage the organization's reputation. Also, IPAB will need to make certain to have appropriate methodologies for calculating cost-effectiveness or comparative effectiveness for each study. For example, they should have good estimates of costs, discount future costs and benefits, and use a better health outcome than five year survival rates (ideally they would use QALYs). Otherwise there will be discrepancies, like between data and the national data.

This case suggests that rationing on the basis of health outcomes with a secondary focus on costs can be politically feasible. Oregon started heavily emphasizing increases in five-year survival rates for the 1992-1994 priority lists and this factor was the most important measure for determining the position on the list [96, 98]. Even though using five-year survival rates is a poor outcome measure because it does not take the quality of life during those years into account, it implies that health improvements are the main consideration during those years. The adjustments that board members made only came as the last step in the process, though they still had a notable effect. This suggests that as long as the IPAB only considers QALYs in their CER studies or at least holds them as its primary concern with costs being secondary it could maintain public support as was the case in Oregon. Though, the federal government had problems with the devaluing of disable people in the rankings and Oregon had to adjust their methods accordingly [98]. This would interfere with CER because these studies often compare treatments using

QALYs, which treat the disabled as having lower QALY values per year of life than non-disabled people.

The rankings on the priority list had weak relationships with cost-effectiveness Oregon. This makes sense given that cost considerations only came in with the adjustments by board members and flaws in the outcome measure of five year survival. The correlation between cost-effectiveness data and the treatment rankings for the priority lists from the years 1992-1994 was positive, meaning that the least cost-effective treatments were more likely to be lower on the list. However, this relationship was not very strong, with correlations less than 0.4. [98]. Thus, if the IPAB makes cost-effectiveness a secondary consideration they could expect this factor to still make a moderate impact. Further, this relationship did not cause public support to waver.

Though, the Oregon example likely represents an overestimate of what the IPAB could feasibly implement. It was not a coincidence that Oregon was the one state that chose this unique method for running their Medicaid program. Instead, it suggests that residents of Oregon have values that are conducive to such a program, whereas citizens of many of the other states lack these values. Since the IPAB would be making national decisions, they would have to be acceptable to people that may not share the same views as the average person in Oregon during this time period. Also, the program actually increased benefits for many conditions from what they were prior to the change to Medicaid [96, 99]. This made it easier for the plan to obtain public support because they did not have to make significant denials of care. Thus, these results may only extend towards the IPAB making decisions that do not amount to serious denials of care.

Section XI. Health Polling Results

Health polling data provides additional information on how the public would likely react to potential policy changes related to IPAB and CER. These findings build upon the feasibility considerations I ascertained from the case studies.

My search for polling information relevant to the topic on the Kaiser Family Foundation database yielded only twelve polls. These polls did not directly mention CER, though they indirectly approached the subject. Thus, the results from these polls relate to general views the public has on CER and the government's role in promoting it.

While these polls suggest that people want there to be better ways of determining the best treatment, they appear wary of the government using this data for policy purposes. In a poll about how much people would trust a federally appointed panel of independent experts that make recommendations on which treatments not to cover, 10% of people would trust this organization greatly, 31% a fair amount, 32% not at all [100]. It would be difficult for such an organization to operate if so many people did not trust its recommendations at all. Another question from the same poll asks the same question without the federal government appointing these experts, and in this case 11% trust the organization a great deal, 44% a fair amount, and 22% not at all. Further, people asked about whether they would support creation of such an organization reported 66% strongly or somewhat in favor and only 31% strongly or somewhat opposed. [100]. So the extent of government involvement decreases the level of trust people have in such an organization. With the new health care legislation in particular, people have had suspicions about the government's motives. For example, more people believe that the new health care legislation would allow a government panel to make decisions for Medicare beneficiaries' end of life care, which the law does not permit [101].

While respondents may consider an independent body making coverage decisions acceptable, people still question the role of science. In a poll on whether there is usually clear scientific evidence on the best treatment for a given situation, only 23% of people answered affirmatively compared with 72% believing that the scientific evidence is not always clear [100]. The wording of this poll does appear misleading since the negative response claims that the scientific evidence *is not always clear*, which does not mean that the evidence does not provide clarity most of the time. Though, the overwhelming support for the latter option suggests that people believe that the current research does not yield sufficient evidence to determine the best treatment in most cases. This result does not necessarily indicate that research will never deliver sufficient evidence to make these determinations. Perhaps there needs to be more comprehensive studies of treatment options before people would believe that the scientific evidence is clear. The literature search on CER did imply that many holes still exist owing to a lack of studies in particular areas. Also, another poll finds that 54% of respondents report that their physicians point to research evidence when justifying their decisions [100]. This implies that evidence matters to both physicians and to the patients they treat.

In summary, relevant, identified health polls suggest public values relating to CER and the IPAB. People also seem open to an independent organization not associated with the government making coverage decisions, while they would have misgivings about the government making these decisions. They also have uncertainty on how much CER can tell us about the advantages of certain treatments.

Limitations:

The strength of these conclusions depends greatly on how these polls were conducted. There are several potential limitations that could limit these conclusions. While they all use

random sampling, the wording of the questions may bias respondents towards certain answers. Also, not everyone has a phone and timing of the calls may eliminate important subpopulations (i.e., day vs. evening or weekday vs. weekend). Confirmation of these results with different wordings would help their credibility. Further, these polls do not directly ask questions about IPAB and its potential use of CER. I had to assume that the sentiments would also apply in that case.

Section XII: Conclusion and Policy Recommendations

In this section, I synthesize my findings to answer my original research question on how the IPAB can promote CER within the Medicare system. First, I explain the primary conclusions I reached from my threefold research. My conclusions address how treatment practices relate to the CER literature and what the case studies and health polling data tells us about political feasibility. Next, I discuss the implications of these conclusions on policy. Specifically, I address requirements for how the IPAB will need to operate so that it can maintain public acceptance and policymaker support. Third, I detail three potential roles for the IPAB in the promotion of CER for Medicare and what the advantages and disadvantages are for each option.

Conclusions from Empirical Section

I reach the following conclusions from my empirical section: 1) variables other than patient preferences and clinical need have associations with treatment decisions, 2) the exact effects of even well-researched treatments are uncertain, 3) Intermountain results do not indicate that watchful waiting results have significantly worse outcomes for prostate cancer treatment, 4) Americans would likely resist government attempts to limit Medicare benefits, and 5) Medicare payments often do not incentivize treatments in accordance with comparative- and cost-

effectiveness data for coverage decisions. I explain each of these conclusions in greater detail and provide justification for each of them below.

Conclusion One:

The first conclusion is that there are indirect decision-making factors affecting selection of treatment. To make this conclusion, I purport that the ideal way for determining treatments involves patients and physicians deciding the best course of treatment given the patient's clinical need, social circumstances, and personal preferences. If physicians' practice systematically differs from the CER literature for particular conditions and there is no reason that patient preferences would drive this change, then this suggests that an indirect or factor external to the patient-provider decision making interaction influences treatment decisions.

I found that factors other than clinical need and patient preferences play a role in the treatment decision. One such example is off-label use of DES, which accounts for a greater proportion of stents than on-label stents. In such cases, there is insufficient evidence showing a clinical advantage for DES [52]. Most likely, patients and physicians extrapolate from the findings for on-label use. In such cases, physicians are probably either misinterpreting the evidence by applying it to off-label use. Such treatment would improve if physicians consider the appropriate CER for the treatment in question. Physicians can provide patients with more accurate pros and cons of the procedures which they are choosing in a shared decision making model.

I also determine that non-evidence based factors affect physician treatment of knee-arthroscopy. The higher rate of TKA suggests that Medicare payments or physician bias may be important factors for treatment choice [61]. The studies in my CER article review suggest that UKA is preferable in many ways to TKA and some studies report UKA as having a greater

improvement in QALYs [57-60]. One potential explanation is that the increased Medicare payment rate for TKAs bias doctors towards this treatment. Physicians may also be unwilling to change their treatment practice because they have done TKAs in the past and are resistant to change. Several studies suggest that physicians resist altering their behavior even in light of new evidence [102, 103]. If either of these accounts is true, then there are extraneous issues that clearly influence treatment decisions.

I also conclude that the review of CER for prostate cancer also indicates indirect variables affecting treatment. The data from my prostate cancer analysis shows the most extreme variation in treatment patterns [64]. However, variation is not a problem if patient preferences or clinical need differs greatly across hospitals and has a high correlation with these treatment differences. However, the distribution of patients with such extremely different needs and preferences does not suggest such a relationship. No explanation can account for why patients in some hospitals do not consider watchful waiting as an option and perceive it as a viable option at another facility without reference to a factor besides preferences and clinical need. It is likely that physicians at different hospitals have a culture that favors certain treatments over others. For instance, physicians at a particular hospital may not believe that watchful waiting is a suitable option for treatment. In such cases, patients make their choices based on their physicians' recommendations and the information they present, which may also have been censored and, therefore, be biased.

I determine that the nationwide dataset on prostatectomies further suggests that seemingly unrelated hospital characteristics affect treatment. The specific variables that had a statistically significant relationship with the rate of prostatectomies include total inpatient days, the ratio of non-teaching to teaching hospitals in a region, the Medicare reimbursement payment

for the last two years of life, the ratio of non-metropolitan to metropolitan hospitals, and the proportion of hospitals with a large number of beds. The significance of these variables suggests that patient preferences and clinical need are not the sole drivers of treatments with external factors to the doctor-patient relationship having an influence on treatment course. That is, the hospital culture may promote prostatectomies depending on whether or not it has the characteristics noted above even without differences in patient attitudes and clinical need.

Conclusion Two:

I find that much uncertainty exists in the CER literature for all the conditions reviewed. After each systematic search, I noted key areas for further study. For example, the long-term impact of the different types of stents on late-thrombosis is not fully understood [51]. Most studies ended too early to obtain accurate, long-term results. Also, some ambiguity remains on the extent of the relative advantages of UKA as compared to TKA when both are viable treatment options [57-60]. The ultimate decision may rest on the patients' preferences, but more clarity would help patients determine which treatment they actually prefer. The most uncertainty I noted was for prostate cancer treatment, as I mention in the first conclusion. The best practice guidelines also expressed the lack of clarity in picking among several options without any one being superior [79, 80]. And the extreme variation in treatment practice implies an arbitrary nature of choice among treatments [64]. Better results on the long-term effects of watchful waiting, in particular, could decrease the large gap in treatment variation if all hospitals recognized watchful waiting as a legitimate treatment option in low-risk cases.

The need for organizations investigating CER to produce or encourage additional research suggests the current literature does not cover all treatments well enough for certainty. NICE conducts many of its own studies because of gaps in the literature where comprehensive

findings are unavailable [104]. Also, the Institute of Medicine (IOM) created a report highlighting the top 100 areas where further CER would help inform treatment decisions. Though, half of this list details aspects of the delivery system. Nevertheless, the IOM could have listed many more conditions that could benefit from CER [105].

Conclusion Three:

I find that watchful waiting does not appear to be significantly detrimental to patients' health in the Intermountain sample. These results serve as a basic CER study. Therefore, the findings do have contributory value for adding to my results from the literature search on prostate cancer. Further, they show the limitations that CER has.

Watchful waiting did not have a statistically significant odds ratio for developing side-effects or a statistically significant odds ratio for requiring active treatment compared to recurrence for other treatments. These results suggest that for the period that this data set covers in a single integrated delivery system, watchful waiting does not lead to either increased odds of developing side-effects or odds of requiring additional treatment. However, the coefficient for the adjusted difference in PSA value was statistically significant, which implies that watchful waiting leads to a significant increase in PSA values. The importance of this measure depends on the how well PSA value predicts the progression of the cancer and whether the adjustment is appropriate. While these conclusions need confirmation with more expansive and long-term studies to address validity concerns, they still present meaningful suggestions of how further studies may contribute to our knowledge base.

Additionally, this study demonstrates common concerns and weaknesses with CER research. One of the limitations was that the study did not cover a sufficiently long time period. While the treatment for prostatectomy may be more effective in the short-term, the long-term

results may not support this finding. I also observed this problem in my review of the literature. For example, the issue with late-thrombosis for heart stents [51]. More long-term studies need to be conducted/funded to overcome this observed limitation in CER studies.

Conclusion Four:

I determine using CER with cost-effectiveness would probably generate resistance against the IPAB. In both the case study on NICE and, to a lesser extent, the Oregon Medicaid experiment, opposition against these programs occurred because of policies that restricted service solely on the basis of CER or similar measurements [85, 89, 90, 95-98]. In particular, the longstanding resistance to the use of cost-effectiveness in making coverage decisions in Medicare may be difficult to overcome, which the introduction details [19, 91]. Oregon had to weaken their attempts at using cost-effectiveness information and instead moved towards using five year survival rates with costs only coming in at the last stage of ranking the treatments [98]. The rankings did have weak to moderate relationships with cost-effectiveness, suggesting that cost-effectiveness could still inform decisions that IPAB makes. Given that public attitudes towards cost-effectiveness in medicine appear low in across the U.S. [19, 91], cost-effectiveness would probably have a limited role similar to the latter years of the Oregon Medicaid system. Using cost-effectiveness in this way would contrast with the key function that economic evaluations have in NICE's decision-making process [85, 87]. However, this difference reflects greater societal acceptance of rationing through CER with cost-effectiveness in the U.K. [91].

Conclusion Five:

I find that that Medicare payment incentives are often inconsistent with the CER evidence. The main problem is that Medicare reimbursement sometimes incentivizes physicians to choose certain treatments that may not provide better outcomes [78]. These situations are

especially problematic when treatments that the payment rates incentivize are actually more expensive than the alternatives. If cheaper treatments could lead to the same health outcomes, less expensive treatment alternatives with similar or better outcomes would be preferable because of the savings in costs to the covered population.

All three conditions had evidence where the literature base did not conclusively prove that a more expensive treatment would significantly improve health outcomes. For heart stents, DES for off-label use did not necessarily result in better health outcomes even though it costs more than BMS [52]. Further, in cases where UKA and TKA both would work, no conclusive evidence shows that one is significantly better across all health measures [57-60]. The reimbursement rate is higher for TKA, which likely induces physicians to do more TKAs than they would if the rates were similar [58]. Finally, doctors receive larger reimbursements for treatments like robotic prostatectomy and IMRT radiation, despite lack of evidence demonstrating improved health outcomes over cheaper alternatives [70, 78]. Also, once physicians purchase a new technology, they have pressure to over utilize it to recover for the initial investment costs [70, 71, 106]. In all these cases, physicians likely face financial incentives to choose one treatment over another despite lack of convincing CER showing improved outcomes from the more expensive treatments.

Policy Recommendations:

In this section, I present policy recommendations based on the conclusions I presented in the previous section. These proposals consist of actions that IPAB could take given the current state of CER in the U.S. That is, with these recommendations, I answer the part of the research question relating to how the IPAB can promote CER given the state of CER and available, relevant examples. The first set of recommendations detail steps the IPAB would need to take no

matter what roles it plans; I refer to these as requirements since failing to fulfill them would undermine the IPAB's ability to encourage the adoption of CER evidence in policy and practice.

Requirement One:

I determine that the IPAB would need to have balanced criteria, monitoring data, and a process for incorporating CER evidence that are open to the public and faultless. Not doing so would lead to even greater controversy. As mentioned as part of the fourth conclusion, the IPAB will likely encounter resistance if it starts having a notable influence on policy. Although some resistance to overt rationing of care (as dictated by application of CER results) will occur no matter what, the critics will have a much better argument if they can point out flaws in the process or approaches to synthesizing the information. If they use a study they did not perform, they need to make sure it is representative of the population they are examining and that the study does not have serious validity concerns. Further, the IPAB need to make sure their approach to gathering and interpreting evidence is sound. Oregon faced much of their initial difficulty because their first methods for calculating the measures for each treatment relied on poor/inadequate data and an inappropriate methodological approach for evaluating the different treatment-condition pairs [98]. Also, NICE has its own detractors, but it has maintained adequate support for their role by focusing on having a well-designed methodology for synthesizing and interpreting results and only using thoroughly investigated studies. NICE also makes its recommendations publicly, so that anyone has access to this information [85, 86, 104, 107]. Thus, NICE's openness and their sound methodological approach to data synthesis and drawing conclusions checks public resistance and provides a rationale for its decisions.

Requirement Two:

I conclude that the policy applications of CER need to primarily focus on the clinical benefit with only a secondary consideration of cost-effectiveness. This relates to the second conclusion in that emphasis on cost-effectiveness would likely lead to protest and would lack political feasibility. Further, talk of CER in the US in the policy context traditionally leaves out the cost-effectiveness criterion [108]. While the U.S. may develop more positive attitudes towards cost-effectiveness in the future, the current atmosphere is not conducive in the public sector. However, cost-effectiveness could have a minor, yet influential, role in the IPAB. For instance, IPAB could encourage the inclusion of cost-effectiveness in the discussions of CER without using it in actual coverage decisions. Such recognition would likely make the information more available to patients and physicians as they make treatment decisions. This availability would influence some patients and physicians to include cost-effectiveness as part of their individual decision-making process. Such information may also be used by health care delivery systems incentivized to improve efficiency and add value in the face of increasing health care costs.

Requirement Three:

I also find that the IPAB should encourage CER to address limitations in literature. This requirement addresses the second, third, and fifth conclusions. Specifically, more CER would help reduce the uncertainty detailed in the second conclusion for the benefits and costs of each treatment. CER evidence will improve the ability of patients and physicians to make informed decisions. Further, such information would help policymakers if they were to actually make coverage decisions using the CER information. One specific treatment option that would benefit from more research is watchful waiting. Specifically, further research on the long-term impacts of condition-specific watchful waiting and in settings other than at Intermountain would help

patients, physicians, and policymakers make decisions about watchful waiting as an option for prostate cancer, in particular. Also, more CER would help clarify when Medicare reimbursements do not correspond well to making all treatments equally attractive for physicians. This would also help policymakers that would want to correct these problematic incentives.

One concern that IPAB should consider when satisfying this requirement is that the studies do cost a non-trivial amount of money. From a governmental or societal-perspective, it may not make sense to research certain conditions if the benefits from increased certainty do not outweigh the costs to perform the study. For this reason the ARRA required the IOM to establish national priorities for CER so that such research investments can be targeted to meet gaps in our knowledge [2].

Roles for the IPAB:

I consider the three most probable roles that the IPAB could take. Each of these roles accord with the conclusions and requirements regarding the use of CER that I determined earlier. The reason for discussing these roles is that they further answer the question of how the IPAB can incorporate CER into the Medicare system. Specifically, this explanation provides the roles that the IPAB may use to promote CER. I will outline these different roles and briefly discuss the advantages and disadvantages of each.

First Role: Alters Payment Rates and Influences Coverage Decisions

In this role, the IPAB could influence Medicare payment rates and the treatments Medicare covers to promote the use of CER rather than enacting across the board payment cuts that would ultimate limit access of beneficiaries, IPAB could use CER evidence to better allocate resources—not paying for treatments that comparatively do not improve quality or those

currently covered that actually cause harm and/or have no effect on patient outcomes. IPAB would, thus, issue recommendations in a manner similar to NICE. However, these recommendations would likely focus on the comparative clinical benefit of treatments instead of the cost-effectiveness given the observations I made around limitations based on cost-effectiveness (conclusion four and requirement two).

The IPAB could recommend no longer covering treatments that CER evidence shows to have worse outcomes or change the payment rate to discourage such options. Further, the IPAB could address situations where two treatments provide the same benefit, but one treatment costs more than the other. They could recommend that Medicare not cover the more expensive treatment, they could alter the payment rate so that physicians would not have an incentive to choose the more expensive treatment, or they could require excess costs be out-of-pocket for the patients. Limiting the role that incentives have on treatment of patients would address the concern I have in the second conclusion that inappropriate factors affect treatment decisions. After IPAB makes these recommendations, Congress would then have to vote them into effect so that ultimate control still rests with elected officials similar to the way NICE operates.

The main advantage of this role would be the extent of impact that it would have on the Medicare system. It makes sense that pricing incentives would greatly influence physicians to adopt treatments in accordance with CER and this finding has support in the literature [109]. Further, it also seems logical that patients would react to price incentives, which also has backing in the literature [110, 111]. Therefore, this option would best address the concern with lack of accordance between Medicare payment rates and the CER literature I note in conclusion five.

The disadvantages of this role relate to political feasibility. Making significant changes to the payment rates would likely meet resistance from physicians, who stand to lose money. Also,

patients and physicians do not want limitations on care, as mentioned in the second conclusion. People would likely resent such forceful promotion of CER and physicians would want all options available to choose from, not just the ones that CER finds acceptable.

Given these considerations, if IPAB came to embody this role, they would have to limit how forceful their changes would be in the promotion of CER. Due to the magnitude of actually altering Medicare reimbursement, this remains the most politically infeasible of the options I consider. Such an option would likely only work under a very favorable political climate that does not currently exist.

Role Two: Clinical Decision Support

The second potential role I consider is for IPAB as a promoter of Clinical Decision Support System (CDSS), which in turn encourages CER. CDSS refers support systems that allow physicians to work with a computer system to guide the diagnostic process and suggest possible treatments given the patient's condition and clinical need [112-115]. The specific CDSS that IPAB would support could then offer an interactive system for patients determining which treatment is most appropriate. The system would offer information on the relevant treatments and vignettes that can help patients by making the information seem more personal, supporting shared decision making. Just presenting patients with information would not fully address their emotional needs. Then, given their physician's advice and their decisions that the CDSS elicits, they can settle on a course of treatment.

The advantages of such a system would be that it preserves and, in fact, enhances the patient's ability to voice their preferences. Dr. Brent James, M.D., M. Stat., a policy expert who has testified before Congress on multiple occasions, identified such a system as one of the two most likely roles that the IPAB could feasibly achieve [116]. Physicians would still have

significant control and would give their full input to the patients. But, the patients also have access to a system that can incorporate the most up to date treatment information and would not have biases related to financial incentives. This system would also limit the impact that extraneous factors have on treatment decisions that I mention in conclusion two. Therefore, this option would probably not incur the public disapproval that I discuss in conclusion two and requirement four. Further, the CDSS system could include information on how costs affect society and the importance of valuing this consideration. Such a reminder at the time of treatment may induce patients to forego more expensive treatments that only provide marginally better or the same outcomes.

The disadvantages of instituting CDSS systems are costs, implementation, and lack of impact. Developing such a system and having them in every hospital would be costly. Further, physicians may not understand how to use such programs, which would make them ineffective for influencing behavior. Further costs would be necessary to ensure that physicians could use these programs. Finally, it is uncertain how much of an impact such a program would have on patient decisions. Anecdotal information for Intermountain Healthcare suggests that shared decision making related to the decision for TKA would reduce elected surgeries by 18 percent [116].

Incentivizing CDSS for shared decision making represents a viable option for IPAB's role. However, more research on the effectiveness of these systems in influencing patient decisions and the costs of installment and implantation would help determine whether IPAB should move in this direction.

Role Three: Moral Suasion

The last potential role for the IPAB is the use of moral suasion to promote CER. IPAB could work with physicians and hospitals to encourage them to consider the CER evidence for treatments and to provide them up-to-date information on new research. Such encouragement could take the form of direct communication, issuing reports, or public announcements that alert hospitals and physicians to what the CER indicates and how they should incorporate it into their practice. The IPAB could also determine the hospitals with practices that do not correspond well to the CER literature and directly target those hospitals and their physicians. Further, the IPAB could include information on cost-effectiveness as part of their moral suasion since they do not actually impose these changes.

The main advantage of such a role for IPAB is political feasibility. This type of system does not directly limit care or the ability for patients and physicians to make decisions, so little objection would exist to its policies of the kind I discuss in conclusion four and requirement two. Also, this is the other type of role that Dr. James considers foreseeable given the political limitations [116]. Another advantage is that the IPAB would require relatively little funding to perform these functions. With this information, physicians would be less likely to make treatment decisions without evidence of how they compare in terms of benefits, regardless of whether they would be reimbursed for more for their services among treatment alternatives. Therefore, this option would address the concern that extraneous factors influence treatment decisions (conclusion two).

The primary disadvantage is that this program would fail to have a large impact on treatment patterns. Hospitals and physicians could ignore the moral suasion from the IPAB. However, most hospital administrators and physicians would likely want to know more about how they could improve the health outcomes of their patients. As for cost-effectiveness, mixed-

evidence exists about whether physicians already consider such information in their treatment decisions despite the incentives to disregard it. One study suggests cost-effectiveness has a small, but measurable impact on treatment practice [117], while another implies that physicians incorporate cost-effectiveness into their analysis and already discuss it somewhat to their patients [118]. Therefore, the IPAB may succeed in getting physicians to consider cost-effectiveness to a greater extent.

Moral suasion remains a practical role for IPAB to take. Further research should look at the kind of impact such an organization could have on treatment decisions and what methods work best before implementing such a program.

In conclusion, these policy recommendations answer how the IPAB can successfully encourage CER within the Medicare system. I made these proposals given the evidence I collected about the current status of the CER literature for the three considered conditions, the two data sets on prostate cancer, the political feasibility of restrictions on care from the two case studies, and the information from the health polls. From these findings, I developed five conclusions. From these conclusions and other sources, I presented three requirements that the IPAB would have to incorporate to operate. Finally, in consideration of these conclusions and the operational necessities, I explained the three most likely roles that the IPAB could assume in order to promote CER in the Medicare system.

Appendix A: Bedsize Categories

Region	Location and Teaching Status	Hospital Bedsize		
		Small	Medium	Large
Northeast	Rural	1-49	50-99	100+
	Urban, nonteaching	1-124	125-199	200+
	Urban, teaching	1-249	250-424	425+
Midwest	Rural	1-29	30-49	50+
	Urban, nonteaching	1-74	75-174	175+
	Urban, teaching	1-249	250-374	375+
South	Rural	1-39	40-74	75+
	Urban, nonteaching	1-99	100-199	200+
	Urban, teaching	1-249	250-449	450+
West	Rural	1-24	25-44	45+
	Urban, nonteaching	1-99	100-174	175+
	Urban, teaching	1-199	200-324	325+

[119]

References

1. Garber, A.M., *Cost-effectiveness and evidence evaluation as criteria for coverage policy*. HEALTH AFFAIRS-MILLWOOD VA THEN BETHESDA MA-, 2004. **23**(4): p. 290-290.
2. Wechsler, J., *Comparative Effectiveness Research May Shape Treatment, Coverage Decisions*. Formulary, 2009. **44**(5): p. 2.
3. Jordan M. VanLare, A.B., Patrick H. Conway, M.D., and Harold C. Sox, M.D., *Five Next Steps for a New National Program for Comparative-Effectiveness Research*. New England Journal of Medicine, 2010. **362**(11): p. 4.
4. Weinstein, M.C. and J.A. Skinner, *Comparative Effectiveness and Health Care Spending-Implications for Reform*. New England Journal of Medicine, 2010.
5. Jacobson, G.A., *Comparative Clinical Effectiveness and Cost-Effectiveness Research: Background, History, and Overview*, C.R. Service, Editor. 2007, Congress: Washington, D.C.
6. Reid, T.R., *The Healing of America. A global quest for better, cheaper and fairer health care*. Journal of Medicine and the Person: p. 1-2.
7. Mushlin, A.I., M.D., and H. Ghomrawi, Ph.D., M.P.H.,, *Health Care Reform and the Need for Comparative-Effectiveness Research*. the New England Journal of Medicine, 2010: p. 4.
8. Stuart Guterman, Karen Davis, and K. Stremikis, *How Health Reform Legislation Will Affect Medicare Beneficiaries*. 2010, Commonwealth Fund.
9. *Patient Protection and Affordable Care Act 2009*.
10. Kaiser Family Foundation, *Summary of New Health Care Reform Law*, in *Focus on Health Reform*. 2010, Kaiser Family Foundation: Menlo Park, CA.
11. Lowes, R., *Drastic Doc Fix Proposed by Federal Debt Commission*.
12. Pearson, S.D. and M.D. Rawlins, *Quality, innovation, and value for money: NICE and the British National Health Service*. JAMA, 2005. **294**(20): p. 2618.
13. Thorpe, K.E., *The rise in health care spending and what to do about it*. Health Affairs, 2005. **24**(6): p. 1436.
14. Cutler, D.M., A.B. Rosen, and S. Vijan, *The value of medical spending in the United States, 1960-2000*. the New England Journal of Medicine, 2006. **355**(9): p. 920.
15. Greenberg, D., et al., *When is Cancer Care Cost-Effective? A Systematic Overview of Cost-Utility Analyses in Oncology*. JNCI Journal of the National Cancer Institute, 2009.
16. Eichler, H.G., et al., *Use of cost-effectiveness analysis in health-care resource allocation decision-making: how are cost-effectiveness thresholds expected to emerge?* Value in Health, 2004. **7**(5): p. 518-528.
17. Rawlins, M.D. and A.J. Culyer, *National Institute for Clinical Excellence and its value judgments*. British Medical Journal, 2004. **329**(7459): p. 224.
18. Neumann, P.J., *Using cost-effectiveness analysis to improve health care: opportunities and barriers*. 2005, New York Oxford University Press.
19. Neumann, P.J., *Why don't Americans use cost-effectiveness analysis*. Am J Manag Care, 2004. **10**(5): p. 308-312.
20. Woolf, S.H., *Potential health and economic consequences of misplaced priorities*. JAMA, 2007. **297**(5): p. 523.

21. Green, L.W., et al., *Diffusion Theory and Knowledge Dissemination, Utilization, and Integration in Public Health*. Annu. Rev. Public Health, 2009. **30**: p. 24.
22. *Reducing the Costs of Poor-Quality Health Care Through Responsible Purchasing Leadership*. 2003, Midwest Business Group on Health.
23. Aaron, H.J. and P.B. Ginsburg, *Is Health Spending Excessive? If so, what can we do About it?* Health Affairs, 2009. **28**(3): p. 16.
24. Davis, K., et al., *Slowing the Growth of US Health Care Expenditures: What Are the Options*. The Commonwealth Fund, 2007.
25. Langreth, R., *Useless Medicine*. Forbes, 2009. **184**(10).
26. Skinner, J.S., D.O. Staiger, and E.S. Fisher, *Is technological change in medicine always worth it? The case of acute myocardial infarction*. Health Affairs, 2006. **25**(2): p. w34.
27. Orszag, P.R. and P. Ellis, *Addressing rising health care costs--a view from the Congressional Budget Office*. the New England Journal of Medicine, 2007. **357**(19): p. 1885.
28. Hobson, K., *Critics Take Aim at High-Tech Care; Many Expensive Medical Tools are Valuable but Arguably Overused*. U.S. News & World Report, 2009. **146**(7).
29. Saha, S., M.D., M.P.H., D.D. Coffman, M.S., and A.K. Smits, M.D., M.P.H., (2010) *Giving Teeth to Comparative-Effectiveness Research--The Oregon Experience*. The New England Journal of Medicine **Volume**,
30. Reinhardt, U.E., P.S. Hussey, and G.F. Anderson, *US health care spending in an international context*. Health Affairs, 2004. **23**(3): p. 10.
31. Sherman, K.L., et al., *Heparin-Coated Stents Do Not Protect Cancer Patients from Cardiac Complications after Noncardiac Surgery*. The American Surgeon, 2009. **75**(1): p. 5.
32. Bill-Axelsson, A., et al., *Radical prostatectomy versus watchful waiting in early prostate cancer*. New England journal of medicine, 2005. **352**(19): p. 1977-1984.
33. Wilson, N.A., et al., *Hip and knee implants: current trends and policy considerations*. Health Affairs, 2008. **27**(6): p. 1587.
34. Zhang, W., et al., *OARSI recommendations for the management of hip and knee osteoarthritis, part I: critical appraisal of existing treatment guidelines and systematic review of current research evidence*. Osteoarthritis and Cartilage, 2007. **15**(9): p. 981-1000.
35. Klein, L.W., *Are drug-eluting stents the preferred treatment for multivessel coronary artery disease?* Journal of the American College of Cardiology, 2006. **47**(1): p. 22.
36. Dawkins, K.D., et al., *Clinical efficacy of polymer-based paclitaxel-eluting stents in the treatment of complex, long coronary artery lesions from a multicenter, randomized trial: support for the use of drug-eluting stents in contemporary clinical practice*. Circulation, 2005. **112**(21): p. 3306.
37. Applegate, R.J., et al., *Comparison of drug-eluting versus bare metal stents on later frequency of acute myocardial infarction and death*. The American journal of cardiology, 2007. **99**(3): p. 333-338.
38. Guinn, B. and S. Keene, *Drug eluting stents vs. bare metal stents: an analysis of stent restenosis and mortality rates. Report*. The Internet Journal of Cardiology, 2008: p. NA.
39. Sigwart, U., *Drug-Eluting Stents Are Safe and Effective: Right or Wrong?* Journal of the American College of Cardiology, 2006. **47**(7): p. 1361.

40. Laarman, G.J., et al., *Paclitaxel-Eluting versus Uncoated Stents in Primary Percutaneous Coronary Intervention*. New England Journal of Medicine, 2006. **355**(11): p. 1105-1113.
41. Spaulding, C., et al., *Sirolimus-eluting versus uncoated stents in acute myocardial infarction*. The New England journal of medicine, 2006. **355**(11): p. 1093.
42. Stone, G.W., et al., *Comparison of a polymer-based paclitaxel-eluting stent with a bare metal stent in patients with complex coronary artery disease: a randomized controlled trial*. JAMA, 2005. **294**(10): p. 1215.
43. Cohen, D.J., et al., *Cost-effectiveness of sirolimus-eluting stents for treatment of complex coronary stenoses: results from the Sirolimus-Eluting Balloon Expandable Stent in the Treatment of Patients With De Novo Native Coronary Artery Lesions (SIRIUS) trial*. Circulation, 2004. **110**(5): p. 508.
44. Shrive, F.M., et al., *Economic evaluation of sirolimus-eluting stents*. Canadian Medical Association Journal, 2005. **172**(3): p. 345.
45. Kaiser, C., et al., *Incremental cost-effectiveness of drug-eluting stents compared with a third-generation bare-metal stent in a real-world setting: randomised Basel Stent Kosten Effektivitäts Trial (BASKET)*. The Lancet, 2005. **366**(9489): p. 921-929.
46. Hofma, S.H., et al., *One year clinical follow up of paclitaxel eluting stents for acute myocardial infarction compared with sirolimus eluting stents*. Heart, 2005. **91**(9): p. 1176.
47. Van Hout, B.A., et al., *One year cost effectiveness of sirolimus eluting stents compared with bare metal stents in the treatment of single native de novo coronary lesions: an analysis from the RAVEL trial*. Heart, 2005. **91**(4): p. 507.
48. Lemos, P.A., et al., *Short-and long-term clinical benefit of sirolimus-eluting stents compared to conventional bare stents for patients with acute myocardial infarction*. Journal of the American College of Cardiology, 2004. **43**(4): p. 704.
49. Lemos, P.A., et al., *Unrestricted utilization of sirolimus-eluting stents compared with conventional bare stent implantation in the "real world": the Rapamycin-Eluting Stent Evaluated At Rotterdam Cardiology Hospital (RESEARCH) registry*. Circulation, 2004. **109**(2): p. 190.
50. Grube, E., et al., *TAXUS I: six-and twelve-month results from a randomized, double-blind trial on a slow-release paclitaxel-eluting stent for de novo coronary lesions*. Circulation, 2002: p. 10000047r.
51. Bavry, A.A., et al., *Late thrombosis of drug-eluting stents: a meta-analysis of randomized clinical trials*. The American journal of medicine, 2006. **119**(12): p. 1056-1061.
52. Bobba, R.K., M. Bollu, and E.L. Arsura, *Off-Label Use of Drug-Eluting Stents*. JAMA, 2007. **298**(8): p. 859.
53. Bavry, A.A., et al., *Development of coronary aneurysm after drug-eluting stent implantation*. Annals of internal medicine, 2007. **146**(3): p. 230.
54. Colombo, A. and S.J. Corbett, *Drug-eluting stent thrombosis: increasingly recognized but too frequently overemphasized*. Journal of the American College of Cardiology, 2006.
55. Kushner, F.G., et al., *FOCUSED UPDATE 2009 Focused Updates: ACC/AHA Guidelines for the Management of Patients With ST-Elevation Myocardial Infarction (Updating the 2004 Guideline and 2007 Focused Update) and ACC/AHA/SCAI Guidelines on Percutaneous Coronary Intervention (Updating the 2005 Guideline and 2007 Focused Update)*. Journal of the American College of Cardiology, 2009. **54**(23).

56. National Institute for Health and Clinical Excellence, *Drug Eluting Stents for the Treatment of Coronary Artery Disease*. 2008: London.
57. Losina, E., et al., *Cost-effectiveness of Total Knee Arthroplasty in the United States*. Archives of Internal Medicine, 2009. **169**(12): p. 1113-1121.
58. Slover, J., et al., *Cost-effectiveness of unicompartmental and total knee arthroplasty in elderly low-demand patients. A Markov decision analysis*. The Journal of Bone and Joint Surgery, 2006. **88**(11): p. 2348.
59. Lombardi, A.V., et al., *Is Recovery Faster for Mobile-bearing Unicompartmental than Total Knee Arthroplasty?* Clinical Orthopaedics and Related Research®, 2009. **467**(6): p. 1450-1457.
60. Xie, F., et al., *Total or partial knee replacement? Cost-utility analysis in patients with knee osteoarthritis based on a 2-year observational study*. EUROPEAN JOURNAL OF HEALTH ECONOMICS. **11**(1): p. 27-34.
61. Mahomed, N.N., et al., *Epidemiology of total knee replacement in the United States Medicare population*. The Journal of Bone and Joint Surgery, 2005. **87**(6): p. 1222.
62. Zhang, W., et al., *OARSI recommendations for the management of hip and knee osteoarthritis, Part II: OARSI evidence-based, expert consensus guidelines*. Osteoarthritis and Cartilage, 2008. **16**(2): p. 137-162.
63. The National Collaborating Centre for Chronic Conditions, *Osteoarthritis: National clinical guideline for care and management in adults*. 2008, Royal College of Physicians.
64. Cooperberg, M.R., J.M. Broering, and P.R. Carroll, *Time Trends and Local Variation in Primary Treatment of Localized Prostate Cancer*. JOURNAL OF CLINICAL ONCOLOGY. **28**(7): p. 1117-1123.
65. London, S., *Adoption of new, costlier therapies for prostate cancer precedes evidence*. NEWS. Internal Medicine News. **43**(7): p. 4.
66. Roddy, E., et al., *Evidence-based recommendations for the role of exercise in the management of osteoarthritis of the hip or knee--the MOVE consensus*. Rheumatology, 2005. **44**(1): p. 67.
67. *Study results from Sidney Kimmel Center for Prostate and Urologic Cancers provide new insights into prostate cancer therapy. Report*. Cancer Weekly, 2010: p. 1032.
68. Lowrance, W.T., et al., *Comparative Effectiveness of Prostate Cancer Surgical Treatments: A Population Based Analysis of Postoperative Outcomes*. JOURNAL OF UROLOGY. **183**(4): p. 1366-1372.
69. Hu, J.C., et al., *Comparative Effectiveness of Minimally Invasive vs Open Radical Prostatectomy*. JAMA: Journal of the American Medical Association, 2009. **302**(14): p. 1557-1564.
70. Lotan, Y., J. Cadeddu, and M.T. Gettman, *The new economics of radical prostatectomy: cost comparison of open, laparoscopic and robot assisted techniques*. The Journal of urology, 2004. **172**(4 Pt 1).
71. Ficarra, V., et al., *Retropubic, laparoscopic, and robot-assisted radical prostatectomy: a systematic review and cumulative analysis of comparative studies*. European urology, 2009. **55**(5): p. 1037-1063.

72. Wei, J.T., et al., *Comprehensive comparison of health-related quality of life after contemporary therapies for localized prostate cancer*. JOURNAL OF CLINICAL ONCOLOGY, 2002. **20**(2): p. 557.
73. Potosky, A.L., et al., *Five-year outcomes after prostatectomy or radiotherapy for prostate cancer: the prostate cancer outcomes study*. JNCI Journal of the National Cancer Institute, 2004. **96**(18): p. 1358.
74. Lim, A.J., et al., *Quality of life: radical prostatectomy versus radiation therapy for prostate cancer*. The Journal of urology, 1995. **154**(4): p. 1420-1425.
75. Potters, L., et al., *Monotherapy for stage T1-T2 prostate cancer: radical prostatectomy, external beam radiotherapy, or permanent seed implantation* 1*. Radiotherapy and Oncology, 2004. **71**(1): p. 29-33.
76. Potosky, A.L., et al., *Quality of life following localized prostate cancer treated initially with androgen deprivation therapy or no therapy*. JNCI Journal of the National Cancer Institute, 2002. **94**(6): p. 430.
77. Bill-Axelsson, A., et al., *Radical prostatectomy versus watchful waiting in localized prostate cancer: the Scandinavian prostate cancer group-4 randomized trial*. JNCI Journal of the National Cancer Institute, 2008.
78. Carreyrou, J. and M. Tamman, *A Device to Kill Cancer, Lift Revenue in The Wall Street Journal*. 2010: New York City, NY.
79. Thompson, I., et al., *Guideline for the management of clinically localized prostate cancer: 2007 update*. Journal of Urology, 2007. **177**(6): p. 2106-2131.
80. National Institute for Health and Clinical Excellence, *Prostate Cancer: Diagnosis and Treatment*. 2008: London.
81. Intermountain Healthcare. *Report to the Community 2009* [cited 4 December, 2010].
82. Intermountain Healthcare. *Awards and Recognition*. 2010 [cited 4 December 2010].
83. Pound, C.R., et al., *Natural history of progression after PSA elevation following radical prostatectomy*. JAMA, 1999. **281**(17): p. 1591.
84. IP Programme, *National Institute for Clinical Excellence*. UK: The clinical effectiveness and cost-effectiveness of new drug treatments for rheumatoid arthritis: etanercept and infliximab, 2001.
85. Schlander, M., *The use of cost-effectiveness by the National Institute for Health and Clinical Excellence (NICE): no (t yet an) exemplar of a deliberative process*. British Medical Journal, 2008. **34**(7): p. 534.
86. Williams, I., S. Bryan, and S. McIver, *How should cost-effectiveness analysis be used in health technology coverage decisions? Evidence from the National Institute for Health and Clinical Excellence approach*. Journal of Health Services Research & Policy, 2007. **12**(2): p. 73.
87. McCabe, C., K. Claxton, and A.J. Culyer, *The NICE cost-effectiveness threshold: what it is and what that means*. Pharmacoeconomics, 2008. **26**(9): p. 733-744.
88. Hirschler, B., *UK's NICE says "no" to four kidney cancer drugs*, in Reuters. 2008: London.
89. Harris, G., *British Balance Benefit vs. Cost of Latest Drugs*, in The New York Times. 2008: New York City, NY.
90. Pfizer, *Final NICE Kidney Cancer Guidance Recommends Sutent(R) (sunitinib) Only*, UK. Medical News Today, 2009.

91. Tunis, S.R., M.D., *Why Medicare Has Not Established Criteria for Coverage Decisions*. The New England Journal of Medicine, 2004. **350**(21): p. 3.
92. *Medicare at a Glance*. 2010, The Henry J. Kaiser Family Foundation.
93. Brekke, K.R. and L. Sørsgard, *Public versus private health care in a national health service*. Health Economics, 2007. **16**(6): p. 579-601.
94. Schlander, M., *Is NICE infallible? A qualitative study of its assessment of treatments for attention-deficit/hyperactivity disorder (ADHD)*. Current Medical Research and Opinion®, 2008. **24**(2): p. 515-535.
95. Dixon, J. and H.G. Welch, *Priority setting: lessons from Oregon*. The Lancet, 1991. **337**(8746): p. 891-894.
96. Bodenheimer, T., *The Oregon Health Plan--Lessons for the Nation--First of Two Parts*. New England Journal of Medicine, 1997. **337**(9): p. 651.
97. Kaplan, R.M., *Value judgment in the Oregon Medicaid experiment*. Medical care, 1994. **32**(10): p. 975-988.
98. Tengs, T.O., et al., *Oregon's Medicaid Ranking and cost-effectiveness: Is there any relationship?* Medical Decision Making, 1996. **16**(2): p. 99.
99. Oberlander, J., T. Marmor, and L. Jacobs, *Rationing medical care: rhetoric and reality in the Oregon Health Plan*. Canadian medical association journal, 2001. **164**(11): p. 1583.
100. Henry J. Kaiser Family Foundation, Harvard University, and NPR The Public And The Health Care Delivery System Survey, *The Public And The Health Care Delivery System Survey*. 2009.
101. Henry J. Kaiser Family Foundation, *Kaiser Health Tracking Poll*. 2010.
102. Mosca, L., et al., *National study of physician awareness and adherence to cardiovascular disease prevention guidelines*. Circulation, 2005. **111**(4): p. 499.
103. Davis, D.A., et al., *Changing physician performance: a systematic review of the effect of continuing medical education strategies*. JAMA, 1995. **274**(9): p. 700.
104. Schlander, M., *Health technology assessments by the National Institute for Health and Clinical Excellence: a qualitative study*. 2007: Springer Verlag.
105. Iglehart, J.K., *Prioritizing comparative-effectiveness research--IOM recommendations*. New England Journal of Medicine, 2009. **361**(4): p. 325.
106. Su, L.M., *Robot-assisted radical prostatectomy: advances since 2005*. CURRENT OPINION IN UROLOGY. **20**(2): p. 130-135.
107. Maynard, A., K. Bloor, and N. Freemantle, *Challenges for the national institute for clinical excellence*. British Medical Journal, 2004. **329**(7459): p. 227.
108. Weinstein, M.C., *Challenges for cost-effectiveness research*. Med Decis Making, 1986. **6**(4): p. 194-198.
109. Rice, T.H., *The impact of changing Medicare reimbursement rates on physician-induced demand*. Medical care, 1983. **21**(8): p. 803.
110. Newhouse, J.P., et al., *Some interim results from a controlled trial of cost sharing in health insurance*. Studying the organisation and delivery of health services: a reader, 2004. **305**: p. 258.
111. Ellis, R.P. and T.G. McGuire, *Supply-side and demand-side cost sharing in health care*. The Journal of Economic Perspectives, 1993. **7**(4): p. 135-151.

112. Charles, C., A. Gafni, and T. Whelan, *Decision-making in the physician-patient encounter: revisiting the shared treatment decision-making model*. Social Science & Medicine, 1999. **49**(5): p. 651-661.
113. Samore, M.H., et al., *Clinical Decision Support and Appropriateness of Antimicrobial Prescribing: A Randomized Trial*. JAMA: Journal of the American Medical Association, 2005. **294**(18): p. 2305-2314.
114. Samore, M.H., et al., *Clinical Decision Support Systems and Antibiotic Prescribing*. JAMA: Journal of the American Medical Association, 2006. **295**(15): p. 1771-1771.
115. Sepucha, K., E. Ozanne, and A.G. Mulley, *Doing the Right Thing: Systems Support for Decision Quality in Cancer Care*. Annals of Behavioral Medicine, 2006. **32**(3): p. 172-178.
116. Dr Brent James, M.D., M.Stat., *Interview with Dr. Brent James, M.D., M.Stat.*, S. Savitz, Editor. 2010: Salt Lake City, UT.
117. Ubel, P.A., et al., *The influence of cost-effectiveness information on physicians' cancer screening recommendations*. Social science & medicine, 2003. **56**(8): p. 1727-1736.
118. Ginsburg, M.E., R.L. Kravitz, and W.A. Sandberg, *A survey of physician attitudes and practices concerning cost-effectiveness in patient care*. Western Journal of Medicine, 2000. **173**(6): p. 390.
119. Agency for Healthcare Quality and Research, *Information on stays in hospitals for participating States from the HCUP State Inpatient Databases (SID) 2010*, HCUPNET.