THE SECOND GENERATION ACCEPTABILITY CURVE: A NOVEL VISUALIZATION APPROACH TO COST-EFFECTIVENESS ANALYSIS

By

Nicholas Gray Micheletti

Thesis
Submitted to the Faculty of the
Graduate School of Vanderbilt University
in partial fulfillment of the requirements
for the degree of

Master of Science
in

Biostatistics

May 10, 2024

Nashville, Tennessee

Approved:

Andrew J. Spieker, Ph.D.

Lauren Samuels, Ph.D.

ACKNOWLEDGMENTS

I would like to thank Andrew Spieker for mentoring, encouraging, and providing advice to me throughout the whole process of writing this dissertation. Both he and Lauren Samuels provided invaluable insight into reviewing and refining this work. I would also like to thank the Vanderbilt University Biostatistics Program for teaching and providing both the resources and people necessary for me to both start and complete this dissertation. Without both my professors, and the wealth of knowledge available at Vanderbilt, I never would have made it this far.

I would also like to thank both my mother and father for always being at my side, and my brothers for encouraging me. My family's support and love motivated me to keep on moving forward in academics and beyond. I love them, and hope to continue doing them proud.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Figure Page

CHAPTER 1

Introduction

In cost-effectiveness analysis, a field dedicated to finding monetarily efficient medical interventions, cost-effectiveness acceptability curves are often used to provide insight into the optimality of one intervention in comparison to another. Such a curve is often calculated using standard p-values (though they are not always acknowledged as such), across a range of possible monetary amounts one could spend. The recent innovation of the second generation p-value, which incorporates null hypothesis intervals to rectify some of the undesirable properties of the standard p-value (namely, inability to distinguish between lack of evidence and evidence of a lack), will enable us to develop a novel improvement to cost-effectiveness visualization methods.

In this work, we will introduce the second generation cost-effectiveness acceptability curve, which uses second generation p-values in lieu of the standard curve's standard p-values. It will be shown that this second generation curve provides improvements upon the first generation; the two most important being the ability to identify a range of values for which two interventions are approximately equivalent in cost-effectiveness, and a range of values for which the results are strictly inconclusive. This innovation upon the standard cost-effectiveness acceptability curve serves to improve comprehension and decision making in cost-effectiveness analysis, for the betterment of the field as a whole.

1.1 Existing approaches for cost-effectiveness analysis

In order to determine which medical strategies are the most efficient from an economically focused cost-benefit perspective, we perform cost-effectiveness analysis (CEA). CEA aids practitioners in allocating limited resources in an informed manner, and helps determine fiscal policy for medical organizations. Due to the relevance of this subject, multiple means of CEA have been developed, as described below.

1.1.1 Incremental cost-effectiveness ratios

A simple and commonly used method of CEA is based on the incremental cost-effectiveness ratio (ICER), which compares the incremental difference in costs and benefits between two medical methods [1]:

$$\text{ICER} = \frac{\text{Mean}(\beta_{x=1}) - \text{Mean}(\beta_{x=0})}{\text{Mean}(\alpha_{x=1}) - \text{Mean}(\alpha_{x=0})} \tag{1.1}$$

Where $x = 1$ and $x = 0$ are two unique interventions, $\beta_{x=1}$ is the cost of intervention $x = 1$, and $\alpha_{x=1}$ is the outcome, or the "effectiveness" of intervention $x = 1$. While the mean is the standard method of calculating an average, there is legitimate and ongoing debate over whether use of the median is more may be more appropriate [2]. Regardless, we will make use of the mean in this work.

There are multiple measures for the effectiveness (β) of an intervention; many involve some means of survival time, whether that be time survived or time spent without specific ailments, but the most common method is measuring quality-adjusted life years (QALYs) [3]. QALYs combine the quality of life and life expectancy into a single variable for easier comprehension. By bootstrapping the sample data and using it to generate multiple ICER's and then creating a confidence interval for them (in order to help quantify uncertainty), the results can be compared to provide insight into the relative effectiveness of two interventions.

1.1.2 Net monetary benefits

While ICER is a useful CEA tool, its methodology fails to account for the reality that resources are finite (and therefore requires the practitioner to set a limit, and search for the best ICER result within that limit). This flaw inspired the use of net monetary benefits (NMB) as a metric [4]:

$$\mu_{x=1} = (\lambda \times \alpha_{x=1}) - \beta_{x=1} \tag{1.2}$$

Where $\mu_{x=1}$ is the NMB of $x = 1$, λ is the set limit of available resources (represented as value in currency), $\alpha_{x=1}$ is the cost of $x = 1$, and $\beta_{x=1}$ is the effectiveness of $x = 1$. If one wants to compare two interventions, they simply calculate the difference of the two NMB's to form the incremental net monetary benefit [4]:

$$\text{INMB} = \mu_{x=1} - \mu_{x=0} \tag{1.3}$$

If the INMB is positive, then intervention $x = 1$ is superior (in terms of cost-effectiveness) to $x = 0$ for the pre-specified λ value; and if it is negative, then $x = 0$ is the superior intervention.

1.1.3 Cost-effectiveness acceptability curves

The cost-effectiveness acceptability curve (CEAc) was created to address the uncertainty involved in CEA (as well as create a better alternative to ICER confidence intervals) [5]. The CEAc uses p-values to compare the INMB's of two interventions over a range of λ values. The p-value for an intervention at a specified λ is found by bootstrapping the sample data, calculating the INMB in each bootstrap sample, and finding the percentage of positive values [5]. Figure 1.1 provides an example CEAc.

Interpretation is intuitive: The x-axis marks the range of possible values for λ, the y-axis marks the 1-sided bootstrapped p-value, and the line color denotes the specific intervention. For example, in figure 1.1, when $\lambda = 0.2$, since intervention $x = 1$ has a p-value of about 1 (and $x = 0$ has a p-value of about 0), it can be reasonably claimed that $x = 1$ is optimal at $\lambda = 0.2$. Further, at the curves' intersection point, the probability of either intervention being

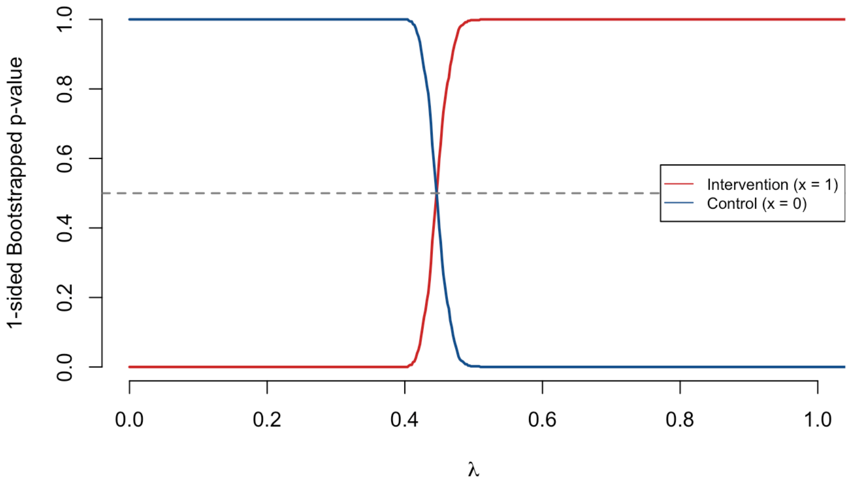

Figure 1.1: An example of the standard cost effectiveness acceptability curve.

optimal is equal (at 50%). This effective modeling of uncertainty through the use of probability found in CEAc's gives them an advantage over simple ICER confidence intervals in decision making by making the probabilistic nature of the analysis clear. Of note, the term "probability" needs to be understood in the context of a given sample space—here, the sample space is over the bootstrapped samples.

1.2 Second generation p-values

P-values have been extensively utilized as a helpful analytical tool (particularly for the purposes of inference and testing hypotheses), but their imperfections have been characterized [6]. Standard p-values can incur an increased type I error rate if repeatedly used (requiring the user to restrict use of the tool, typically to a single pre-specified analysis that is meticulously adhered to), can't be used to make inferences supporting the null hypothesis, and can be easily misused to create erroneous results (e.g. "p-hacking"). The second generation p-value was created in order to address these issues, and in doing so create a alternative to the standard p-value that can be seen as superior in these aspects [7].

The main innovation of the second generation p-value is its use of an interval null hypothesis, as compared to the point null hypothesis that a standard p-value uses. Let H_0 be the pre-specified null hypothesis interval and I_c be an interval of hypotheses that properly represents a scalar variable c (typically, although not exclusively, confidence intervals covering c are used for I_c). Then the second generation p-value, p_δ, can be calculated as follows [7]:

$$p_\delta = \frac{|I_c \cap H_0|}{|I_c|} \times \max\left\{\frac{|I_c|}{2|H_0|}, 1\right\} \tag{1.4}$$

Since the second generation p-value is simply the ratio of the length of the intersection between I_c and H_0 to the length of I_c, multiplied by a "correction factor" ($\max\{|I_c|/2|H_0|, 1\}$), its interpretation is different than that of a standard p-value [7]. Namely, p_δ represents the ratio of hypotheses that are null hypotheses, so if $p_\delta = 0$ then the data fully supports the alternate hypotheses, while if $p_\delta = 1$ the data fully supports the null hypotheses, and if $p_\delta = 0.5$ by any means then the results are wholly inconclusive. In practice, unless p_δ is approximately 0 or 1, then the result is typically interpreted as inconclusive [6].

The second generation p-value avoids many of the problems faced by the standard p-value. By utilizing intervals for H_0, the type I error rate can be reduced. Scientific inferences in support of the null hypothesis interval can be made (if supported by the data), just as they can for an alternate hypothesis. Finally, forcing researchers to pre-specify their null hypothesis intervals prevents them from falling into one of the pitfalls seen in erroneous standard p-value calculation: changing the null hypothesis to produce more favorable results. These benefits can only be realized however, if I_c (or rather, the process for calculating I_c) and H_0 are properly specified before a study starts. If this pre-specification does not occur, one might erroneously choose intervals based on their desired conclusion. With proper pre-specification of I_c and H_0, second generation p-values prove to be an overall superior alternative to the standard p-value. Using second generation p-values in the place of standard p-values could produce improvements in other analytical methods, especially cost-effectiveness acceptability curves.

CHAPTER 2

Methodology

2.1 Constructing a second generation cost-effectiveness analysis curve

The second generation cost-effectiveness acceptability curve (2CEAc) improves upon the standard CEAc by calculating the degree of overlap between the cost-effectiveness interval estimate (generated from incremental net monetary benefit confidence intervals) of an intervention and an interval null represented by the second generation p-value, as opposed to the standard's basis on the 1st-generation p-value. The main benefit of implementing of this method is the improved ability in making meaningful conclusions if the data supports the null hypothesis interval, and a better characterization of the cost-effectiveness of interventions. This necessitates the calculation of the second generation p-value, determining graphical depiction, and showing how the 2CEAc is interpreted as compared to a standard CEAc.

2.1.1 Finding the second generation p-value at a λ value

In order to construct a 2CEAc, a sample data-set from a larger population containing two interventions' costs and effects must be supplied. Let $x = 1$ and $x = 0$ be two distinct interventions, β_1 and β_0 be the monetary costs of $x = 1$ and $x = 0$ respectively, and α_1 and α_0 the "effectiveness" of $x = 1$ and $x = 0$ respectively. Note that as explained in Section 1.1, there are various measures of effectiveness; the measure picked should fit with contemporary standards. It would also be helpful for the practitioner to select an acceptable range of resources that are acceptable to be used, or in other words a range of λ_j values represented as Λ. For this paper, all 2CEAc's will have Λ's individual λ_j values separated by a thousandth of the value range being analyzed, meaning there will be 1001 unique λ_j in Λ. As an example: if Λ ranges from 1 to 2 million dollars, then the values of λ would be marked in the thousands. More λ_j's in Λ could be used if the practitioner felt it necessary, the only limitation being the time consumed in calculation.

Another important value must be decided: the null hypothesis interval of incremental net monetary benefit (INMB) values, represented as H_0. A good interpretation of H_0 is as an interval of values representing how different the cost-effectiveness estimates of interventions have to be from each other before being considered substantively different. If the range is kept too small in an attempt to find minute differences, then the benefits of the second generation p-value's interval null hypothesis aren't being properly utilized, creating a graph that looks very similar to a standard CEAc (whose null hypothesis is a point instead of an interval). If the null hypothesis interval range is made too large in an attempt to only identify large differences in cost-effectiveness between interventions, then the 2CEAc fails to capture ranges of λ values for which one intervention is optimal. Unless otherwise specified, a null hypothesis interval of $(-0.15 \times \max(\Lambda), 0.15 \times \max(\Lambda))$ will be used in examples. One should note that all null hypothesis interval lengths should be symmetric, in order that the 2CEAc is not predisposed toward either intervention.

After specifying the null hypothesis interval for the 2CEAc, the level of the confidence intervals generated should also be selected. The level of a confidence interval, whether it is 90%, 95%, or 99%, can end up changing the calculating second generation p-value. Choosing the level of a confidence interval after conducting exploratory analyses can tempt the practitioner to choose an interval that returns desired second generation p-values. For all 2CEAc's created in this paper, a 95% confidence interval level is pre-specified.

The first step in creating the 2CEAc is to bootstrap the data and calculate the INMB's for each bootstrapped sample. Let $x = 1$ and $x = 0$ represent two unique interventions for which data is collected. Take the data and re-sample it with replacement (so the same data points can be sampled multiple times). INMB's need to be calculated for each λ_j value, and since $\lambda_j \in \Lambda$, Λ will be used for calculation. Calculate the INMB's for this re-sampled (or bootstrapped) data by combining equations (1.2) and (1.3):

$$\text{INMB}_i = (\Lambda \times \alpha_{1i}) - \beta_{1i} - ((\Lambda \times \alpha_{0i}) - \beta_{0i}) = \Lambda(\alpha_{1i} - \alpha_{0i}) - (\beta_{1i} - \beta_{0i}) \tag{2.1}$$

Where i specifies the specific re-sampled data-set being used, so β_{1i} represents the monetary cost of intervention $x = 1$ based on the re-sampled data of the i'th iteration, for example. Each INMB_i is a vector with as many values as $\lambda_j \in \Lambda$, each of which can be specified as INMB_{ij}. Let B denote the number of bootstrap replicates. Making $B \geq 500$ often creates satisfactory and consistent results, but there are studies that warrant more bootstraps based on the nature of the data. Next, for each individual λ_j value, calculate a 95% confidence interval of values, represented by INMBCI_j. This can be performed through a variety of bootstrap confidence interval methods. For example, a studentized interval can be found by calculating the t-statistic of INMB_{ij} for each individual i, and then calculating the 2.5% and 97.5% quantile of the results. Another example is the jittered bootstrap, which applies a smoothing function to find an estimate of INMB_{ij} for each individual i, and then calculates the 2.5% and 97.5% quantile of the results. This paper will use the quantile-based method, for its simplicity and versatility in use. The method is done by simply finding the 2.5% and 97.5% quantile of INMB_{ij} for each individual i.

With the collection of INMBCI_j's comprising INMBCI, and the null hypothesis interval H_0 justified, the second generation p-value can be calculated. Inputting these values into the formula for second generation p-values (1.4), the second generation p-value is then given by:

$$p_{\delta_{\lambda_j}} = \frac{|\text{INMBCI}_j \cap H_0|}{|\text{INMBCI}_j|} \times \max\left\{\frac{|\text{INMBCI}_j|}{2|H_0|}, 1\right\} \tag{2.2}$$

Where $p_{\delta_{\lambda_j}}$ is the second generation p-value for a specified λ_j value. $|\text{INMBCI}_j \cap H_0|$ can be calculated by creating an interval whose lower bound is the maximum of INMBCI_j's and H_0's lower bounds and whose upper bound is the minimum of INMBCI_j's and H_0's upper bounds.

With the collection of all $p_{\delta_{\lambda_j}}$'s, written as p_δ, having been calculated, all necessary materials for constructing a 2CEAc have been acquired.

2.1.2 Depicting the curve

Since p_δ is calculated in such a way that a value of 1 means that the data supports the interval null hypothesis, and 0 means that the data supports an alternate hypothesis, for the purpose of intuitive comprehension the y-axis will be $1 - p_\delta$ instead of p_δ. It should be noted that the strength of the support for both the null and alternate hypotheses is proportional to the the percentage of the confidence interval generated. For $1 - p_\delta$ then, a value of 1 means that the data supports only the alternate hypothesis, and 0 means that the data supports only the interval null hypothesis.

While it might be tempting to automatically graph the relationship between the λ and $1 - p_\delta$ values as in a standard CEAc, doing so will create an unsatisfactory result, as seen in Figure 2.1. This curve fails to tell us the most important piece of information we want to know: that is, the circumstances under which which each intervention optimal. The reason why Figure 2.1 has this flaw is inherent in the way the null hypothesis interval is specified. Namely, the null hypothesis being that there is no meaningful difference between the two interventions in cost-effectiveness. The alternate hypothesis therefore is that there is a meaningful difference between the two interventions in cost-effectiveness. It is clear some method of distinguishing interventions graphically is necessary, in order to account for the alternate hypothesis.

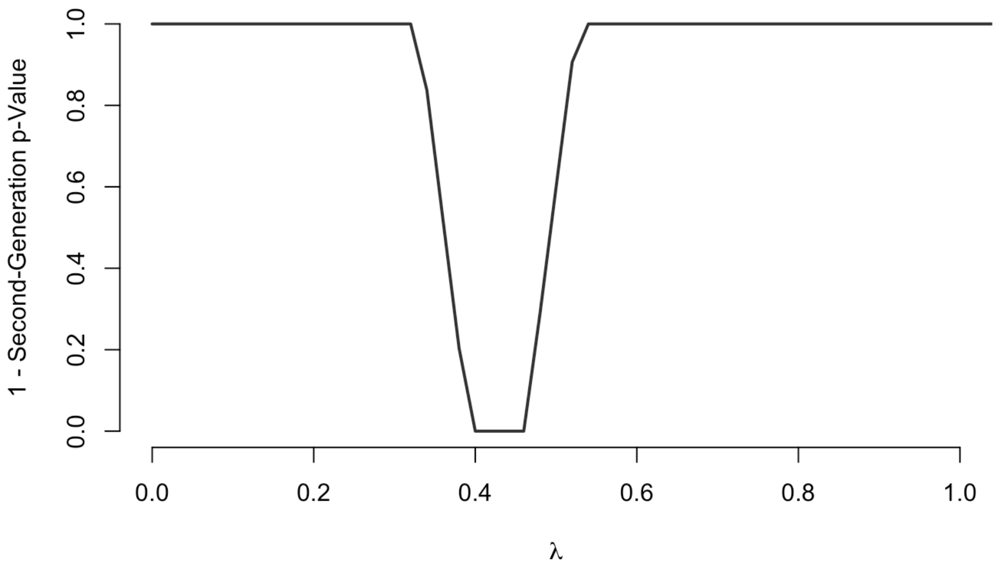

Figure 2.1: A rudimentary example second generation cost effectiveness analysis curve.

We consider two methods for distinguishing between interventions: an overlapping method, and a method using confidence intervals. Instead of defining the null hypothesis interval as there being no significant difference between the two interventions in cost-effectiveness, let the null hypothesis interval describe one specified intervention not being optimal over the other. Thus, let $H_0 : (-\infty, 0.15 \times \max(\Lambda))$, or rather that intervention $x = 1$ is not optimal in terms of cost-effectiveness. Using this new H_0, and repeating the process described in 2.1.1, the 2CEAc shown in Figure 2.2 is created.

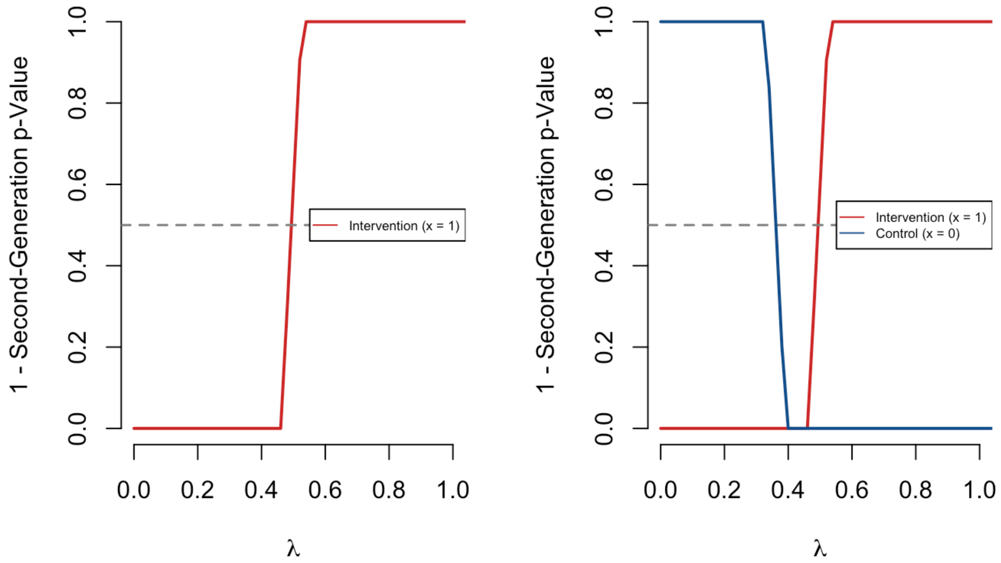

Figure 2.2: A rudimentary example second generation cost effectiveness analysis curve for individual interventions $x = 1$ and $x = 0$.

If the practitioner is only interested in the optimality of one of the two interventions, then the left graph in Figure 2.2 is a sufficient result. If there is interest in both interventions, however, this graph is insufficient. Making $H_0 : (-0.15 \times \max(\Lambda), \infty)$, or rather that $x = 0$ is not the optimal intervention in terms of cost-effectiveness, and repeating the process described in 2.1.1, and overlaying that 2CEAc describing $x = 0$ with the 2CEAc describing $x = 1$, the right-hand graph of Figure 2.2 is produced. Removing the lines where $1 - p_\delta = 0$ for only $x = 1$ or only $x = 0$, the generated 2CEAc looks identical in shape to Figure 2.1.

Since the intersection of the 2 H_0's $(-0.15 \times \max(\Lambda), \infty)$ and $(-\infty, 0.15 \times \max(\Lambda))$ is $(-0.15 \times \max(\Lambda), 0.15 \times \max(\Lambda))$, the base H_0 that was being used in Figure 2.1, this graph is the 2CEAc made in Figure 2.1 with interventions identified. This is the method of calculating two null hypothesis intervals and combing their results graphically, which is called the overlapping method.

While the overlapping method creates functional 2CEAc's, it requires twice the amount the calculations made in Section 2.1.1. A second method, the confidence interval method, can be used instead to achieve the same result with fewer mathematical processes required. If the INMBCI_j has a smaller lower bound than H_0, then the 2CEAc should be marked as favoring intervention $x = 0$ at that λ_j value. If the INMBCI_j has a larger upper bound than H_0, then the 2CEAc should be marked as favoring intervention $x = 1$ at that λ_j value. In the event that INMBCI_j has both a smaller lower bound and larger upper bound than H_0, whichever bound is farther from H_0 designates the favored intervention.

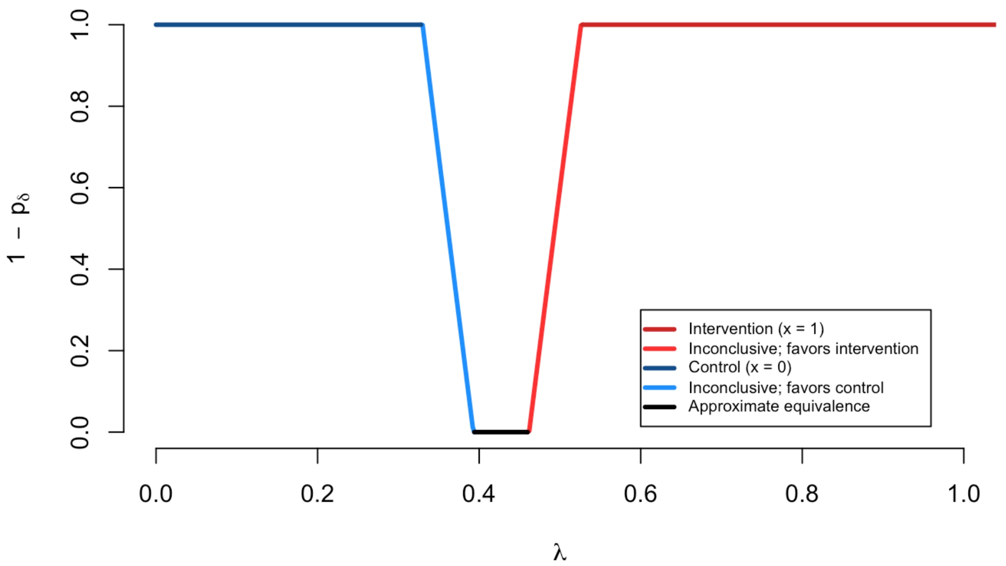

Figure 2.3: An example second generation cost effectiveness analysis curve using the confidence interval method.

So if there is a larger range between the lower bounds than the upper bounds, then the λ_j value is marked as belonging to $x = 0$. Further, λ vales where $p_\delta = 0$ are marked as showing approximate equivalence between interventions. This creates Figure 2.3, which looks identical to the left graph of Figure 2.2.

Since Figure 2.3 and 2.2 are identical in intervention designation, and figure 2.2 was shown to be an accurate 2CEAc depiction, it follows that the confidence interval method is valid. This makes sense, as Figure 2.3's method of comparing the confidence interval and null hypothesis lengths is how second generation p-values are calculated. All 2CEAc's shown for the rest of the paper utilize this confidence interval process. With this, all necessary information for a 2CEAc is shown, and all that remains is interpreting the results.

2.1.3 Interpreting the curve

A 2CEAc cannot be interpreted in the same way one would interpret a standard CEAc, due to the properties held by the second generation p-value. Using the explanation in Section 1.2 on second generation p-values, $1 - p_\delta$ can be defined as the ratio of alternate hypotheses to null hypotheses that are supported by the gathered data [7]. Therefore, if $1 - p_\delta = 0$ the data supports only null hypotheses, and if $1 - p_\delta = 1$ the data supports only alternate hypotheses.

Based upon this information, if $1 - p_\delta \approx 0$ on the 2CEAc, then one can claim that the data supports the hypothesis that neither intervention is strictly optimal at that λ value. Conversely, if $1 - p_\delta \approx 1$ on the 2CEAc, then the statistician can claim that the data supports the hypothesis that there is an optimal intervention at that λ value. Specifically, the color of the 2CEAc at that λ value would designate the optimal intervention.

If $0 < 1 - p_\delta < 1$, then a more developed explanation is required. When $1 - p_\delta = \frac{1}{2}$, then we claim that the results are altogether inconclusive. That is, neither the null or alternate hypotheses has greater support from the data. The closer $1 - p_\delta$ is to $\frac{1}{2}$, the greater the level of how inconclusive the result is. Therefore, if $0 < 1 - p_\delta < 0.5$, then the statistician can claim that the data somewhat supports the hypothesis that neither intervention is strictly optimal at that λ value, with a level of inconclusiveness inversely related to the distance from 0.5. Similarly, if $0.5 < 1 - p_\delta < 1$, then the statistician can claim that the data somewhat supports the hypothesis that a specified intervention (by whichever color the 2CEAc line is at that point) is strictly optimal at that λ value, with a level of inconclusiveness inversely related to the distance from 0.5. As an example, if $1 - p_\delta = 0.6$, then the alternate hypothesis is supported with a high level of inconclusiveness, and if $1 - p_\delta = 0.9$, then the alternate hypothesis is supported with a low level of inconclusiveness.

Taking what has been stated, the method of interpretation can be summarized as thus:

1. If $1 - p_\delta \approx 0$, then the data supports the hypothesis that neither intervention is optimal at that λ value.

2. If $1 - p_\delta \approx 1$, then the data supports the hypothesis that the specified intervention is optimal at that λ value.

3. If $0 < 1 - p_\delta < 0.5$, then the data somewhat supports the hypothesis that neither intervention is optimal at that λ value, with a level of inconclusiveness inversely related to the distance from 0.5.

4. If $0.5 < 1 - p_\delta < 1$, then the data somewhat supports the hypothesis that the specified intervention is optimal at that λ value, with a level of inconclusiveness inversely related to the distance from 0.5.

5. If $1 - p_\delta \approx 0.5$, then the results are inconclusive at that λ value.

2CEAc interpretation can also be better understood by associating it with a graph comparing λ and $INMB_\lambda$, as shown in figures 2.4 and 2.5, with a 95% confidence interval included. Each λ vs. $INMB_\lambda$ plot contains a specified null hypothesis centered on zero, under varying conditions. When the confidence interval is completely below the null hypothesis, the 2CEAc considers the control ($x = 0$) optimal, and when it is completely above, the treatment of interest ($x = 1$) is optimal. If the confidence interval is contained completely within the null, then the 2CEAc considers the treatments equivalent in cost-effectiveness. This is shown in “wide interval” graphs, a scenario similar to what was discussed in Figures 2.1 to 2.3. The proportion of the confidence interval outside the null hypothesis interval is what is being depicted in the 2CEAc. If the confidence interval’s range contains an area both inside and outside the null hypothesis interval, then the intervention favored by the confidence interval (that is, the intervention with a larger range of values covered) is marked in the 2CEAc by that intervention’s color.

In the event that the ratio of coverage of the confidence interval that exists within the null hypothesis interval is 0.5, as seen in the “narrow interval (which is analogous to the “wide interval” scenario, but with a smaller null hypothesis interval), then the 2CEAc considers the results strictly inconclusive at that λ value. This is why it is important to set a reasonable null hypothesis interval, to prevent unnecessary inconclusively. “Widening CI” marks a scenario seen in

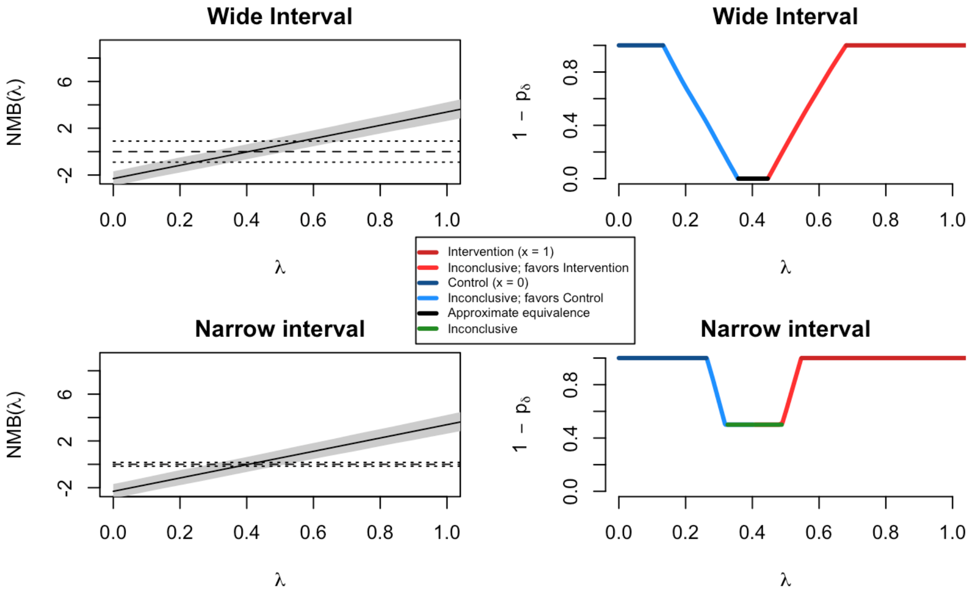

Figure 2.4: NMB vs. λ plots, as compared to 2CEAc's: wide and narrow null intervals marked by dotted lines.

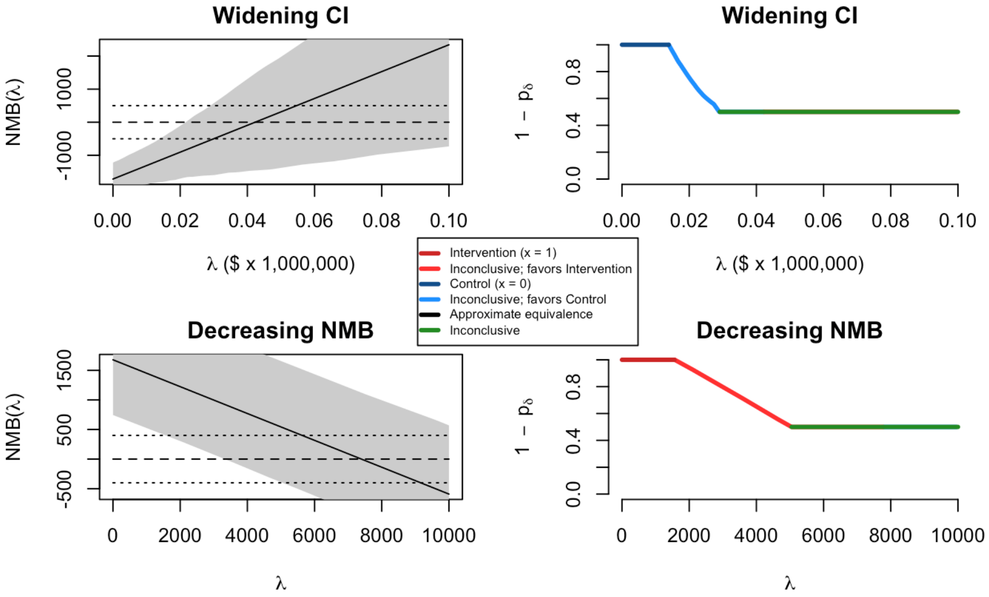

Figure 2.5: NMB vs. λ plots, as compared to 2CEAc's: widening and decreasing confience intervals over null intervals marked by dotted lines.

cases where either the data spread is extremely high, or the amount of data at hand is low: a confidence interval that widens as λ increases. When the confidence interval contains the entire null hypothesis, alongside areas above and below it, the result is naturally inconclusive. Finally, “Decreasing NMB” is presented to better help intuition of the 2CEAc in a different scenario, where the treatment of interest ($x = 1$) is inferior in cost-effectiveness as compared to the control ($x = 0$) as $\lambda \to \infty$. Since $INMB_\lambda$ starts above the null hypothesis interval, the 2CEAc starts favoring the treatment of interest. Understanding this relationship between confidence interval and null interval coverage is key to interpreting the 2CEAc.

CHAPTER 3

Simulations

3.1 Simulation setup

In order to better understand the 2CEAc, multiple simulation studies will be conducted. Sections 3.1.1-3.1.3 will explore the relationship between changes in data and curve shapes, specifically looking at how changes in costs, effects, data spreads, and interval null hypothesis lengths affect the shape of the 2CEAc. Sections 3.1.4-3.1.6 will compare the standard CEAc to the 2CEAc, and showcase how the 2CEAc outperforms the CEAc in certain situations.

All simulations of costs and effectiveness given will be conducted under similar normal distributions in order to better demonstrate the aspects of the 2CEAc. For each simulation, interventions $x = 1$ and $x = 0$ will have $\frac{N}{2}$ samples taken, where N is the total number of samples taken for both interventions, and 500 bootstraps will be taken of the data. $\alpha_{x=0} = N(50, \sigma)$, $\alpha_{x=1} = N(50+\mu, \sigma)$, $\beta_{x=0} = N(100, 40\sigma)$, $\beta_{x=1} = N(100+\gamma), 40\sigma)$. Λ will consist of λ values ranging from $\{0, 100\}$. The null interval will be $H_0 : \{-0.15 \times 100, 0.15 \times 100\} = \{-15, 15\}$, unless otherwise specified.

All simulations were created using R code. Similarly, all generated standard CEAcs and 2CEAcs were created using *R*. An *R* code outline can be found at the bottom of the paper, in Chapter 6.

3.1.1 Simulation A: exploring how changes in cost and effects affect curve shape

Let $\sigma = 4$, $N = 20000$. Four 2CEAcs will be generated, each with a unique μ and γ combination (so the only difference between graphs will be the differences in the $\alpha_{x=1}$ and $\beta_{x=1}$ values). The four graphs will consist of $\mu = 4$ $\gamma = 160$; $\mu = 4$ $\gamma = 128$; $\mu = 3.2$ $\gamma = 160$; and $\mu = 3.2$ $\gamma = 128$, in order.

3.1.2 Simulation B: exploring how changes in data spread affect curve shape

Let $\mu = 4$, $\gamma = 160$, $N = 20000$. Four 2CEAcs will be generated, each with a unique σ value. The four graphs will consist of $\sigma = 4$; $\sigma = 8$; $\sigma = 12$; $\sigma = 20$, in order.

3.1.3 Simulation C: exploring how changes in null interval lengths affect curve shape

Let $\sigma = 4$, $\mu = 4$, $\gamma = 160$, $N = 20000$. Four 2CEAcs will be generated, each with a unique H_0 length. As a reminder, $H_0 : \{\text{interval value} \times \max(\Lambda)\}$, where $\max(\Lambda) = 100$. The four graphs will consist of $H_0 : \{-0.1 \times 100, 0.1 \times 100\} = \{-10, 10\}$; $H_0 : \{-0.15 \times 100, 0.15 \times 100\} = \{-15, 15\}$; $H_0 : \{-0.2 \times 100, 0.2 \times 100\} = \{-20, 20\}$; $H_0 : \{-0.4 \times 100, 0.4 \times 100\} = \{-40, 40\}$, in order.

3.1.4 Simulation D: comparing the standard and second generation curves with a low variance, high sample size example

Let $\sigma = 4$, $\mu = 4$, $\gamma = 160$ $N = 20000$. $\alpha_{x=0} = N(50,4)$, $\alpha_{x=1} = N(50+4,4)$, $\beta_{x=0} = N(100,160)$, $\beta_{x=1} = N(100+160,160)$. Both the standard CEAc and the 2CEAc will be generated from this data.

3.1.5 Simulation E: comparing the standard and second generation curves with high data variance

Let $\sigma = 20$, $\mu = 4$, $\gamma = 160$ $N = 20000$. $\alpha_{x=0} = N(50,20)$, $\alpha_{x=1} = N(50+4,20)$, $\beta_{x=0} = N(100,800)$, $\beta_{x=1} = N(100+160,800)$. Both the standard CEAc and the 2CEAc will be generated from this data.

3.1.6 Simulation F: comparing the standard and second generation curves with a low sample size

Let $\sigma = 20$, $\mu = 4$, $\gamma = 160$, $N = 800$ (so each intervention has 400 samples taken). $\alpha_{x=0} = N(50,4)$, $\alpha_{x=1} = N(50+4,4)$, $\beta_{x=0} = N(100,160)$, $\beta_{x=1} = N(100+160,160)$. Both the standard CEAc and the 2CEAc will be generated from this data.

3.2 Simulation results

Note that while $\Lambda \in \{0,100\}$, for the sake of better visualization, all 2CEAcs and standard CEAcs show only λ values ranging from 30 to 60. Figures 3.1 to 3.6 correlate to sections 3.1.1 to 3.1.6 respectively, so Figure 3.3 is the results for 3.1.3: simulation C for example.

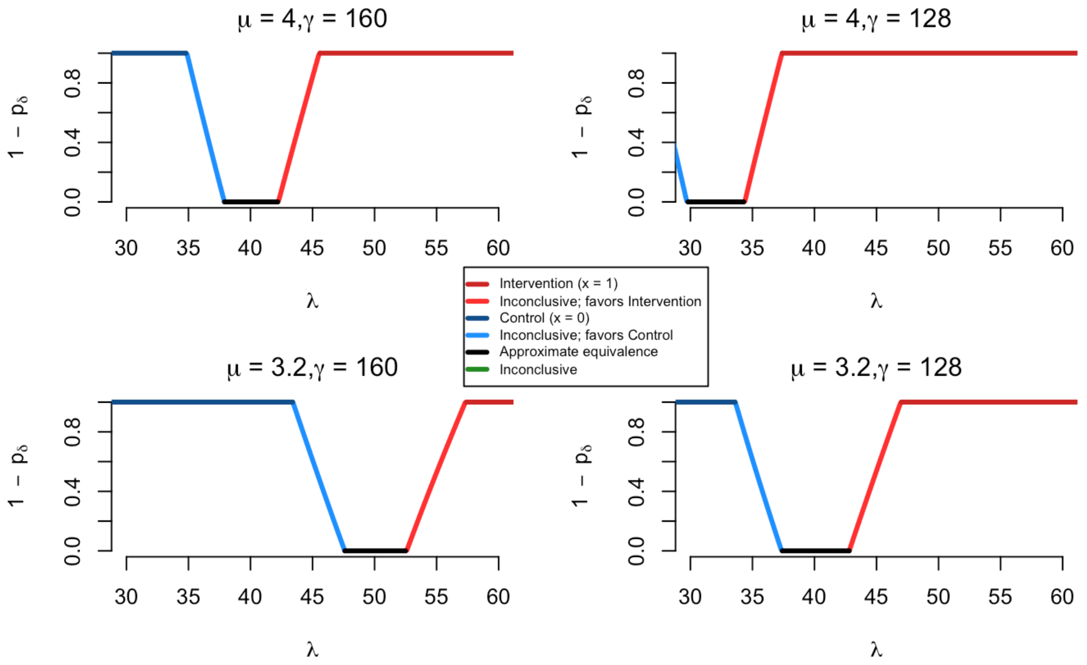

Figure 3.1: Simulation A: second generation cost effectiveness analysis curves where μ, γ values differ.

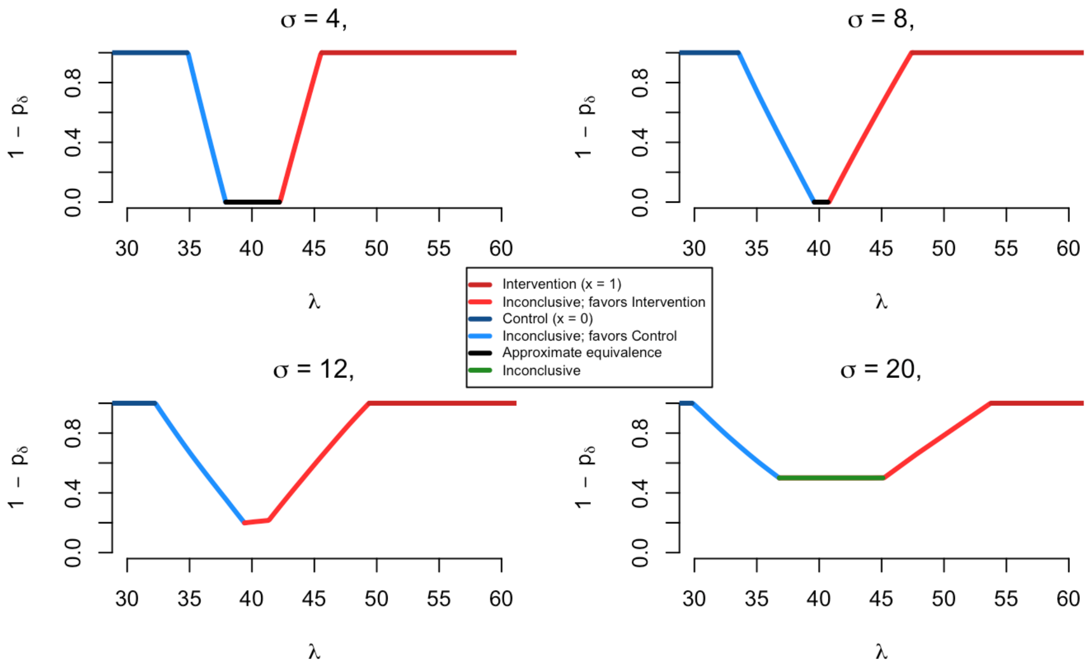

Figure 3.2: Simulation B: second generation cost effectiveness analysis curves where σ values differ.

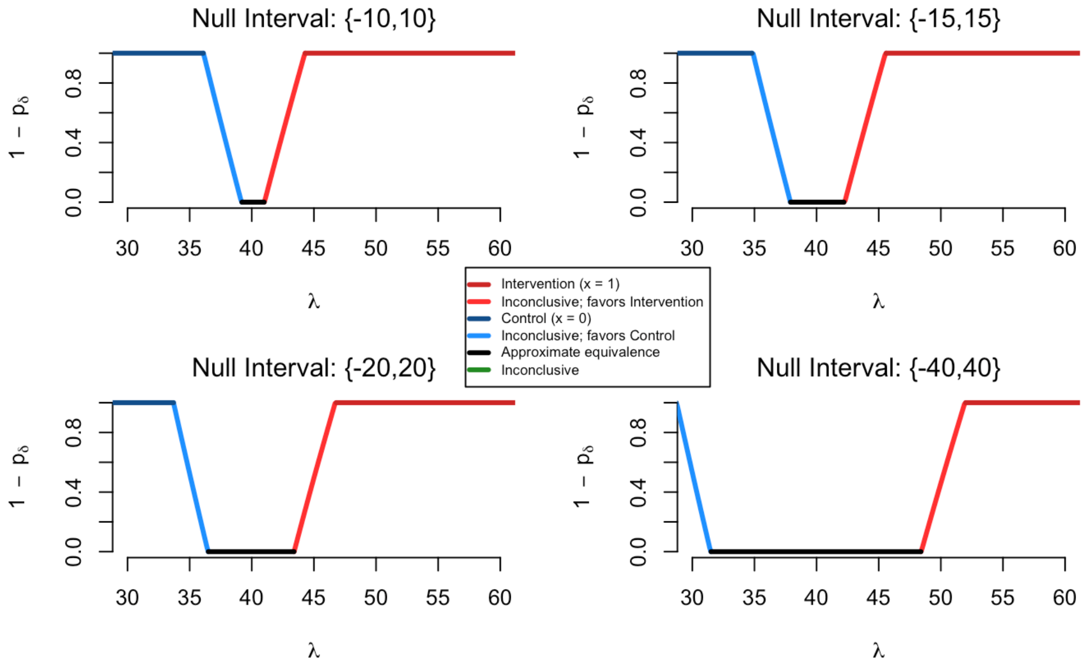

Figure 3.3: Simulation C: second generation cost effectiveness analysis curves where H_0 lengths differ.

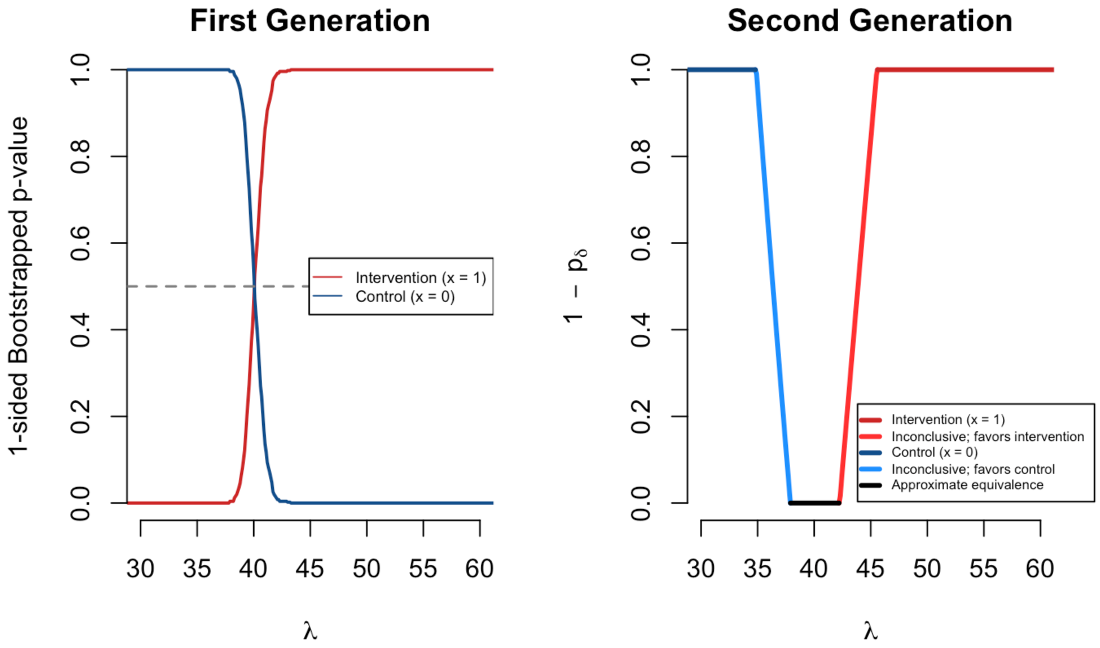

Figure 3.4: Simulation D: standard vs. second generation cost effectiveness analysis curves where $\sigma = 4$.

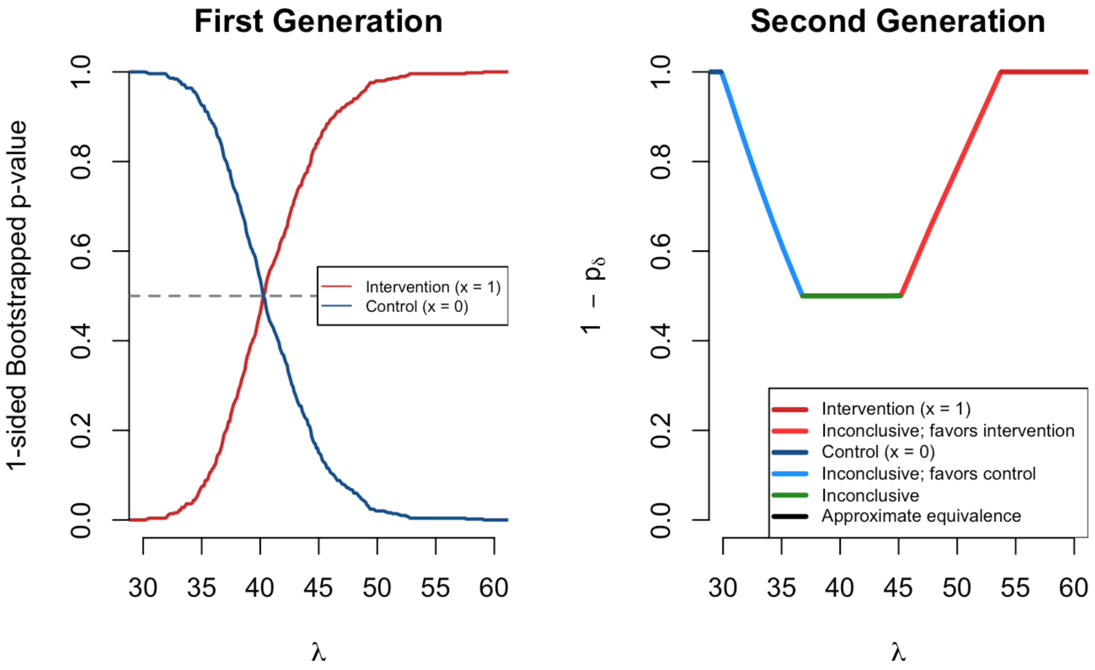

Figure 3.5: Simulation E: standard vs. second generation cost effectiveness analysis curves where $\sigma = 20$.

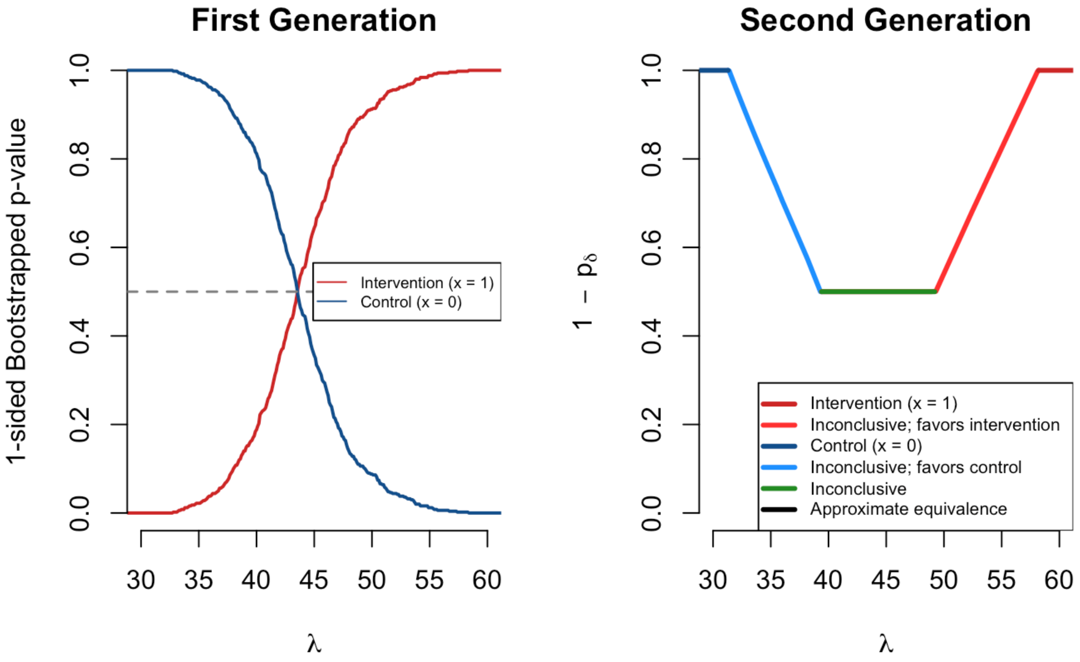

Figure 3.6: Simulation F: standard vs. second generation cost effectiveness analysis curves where $\sigma = 4$, $N = 800$.

CHAPTER 4

Application

4.1 Cancer treatment application

Beyond the theoretical examples posed in Section 2.2, the benefits of using a 2CEAc can be demonstrated effectively by employing it on an applied example. In *Barr et al. (2023)*, the authors focus on new cancer treatments, and the statistical uncertainty their datasets sometimes bring [8]. Cancer treatments can range wildly in costs and effects in terms of both average values and overall spreads. Further, with new treatments having higher costs, long treatment times, and challenging data-gathering requirements, sample sizes tend to be low. Combining these traits together, studies of new interventions in the field of cancer treatment tend to hold a high amount of uncertainty. The paper discusses and analyzes different cost-effectiveness techniques in terms of how they incorporate and depict uncertainty for future analytical work in the field of cancer. One such analytical method covered is the standard CEAc. The main conclusion put forth by the authors of the paper is that while the standard CEAc is useful, it fails to represent uncertainty in any meaningful capacity.

This is understandable, as the standard CEAc does not directly depict inconclusivity; instead it incorporates it into its bootstrapped 1-sided p-values, interpreted as the probabilities of an intervention being optimal. It's hard to draw a meaningful conclusion on inconclusiveness when the probabilities are below 0.9 or above 0.1, for either intervention. As discussed previously, however, the 2CEAc rectifies this issue. This section will run a similar simulated dataset as the one constructed in *Barr et al. (2023)* over various null interval lengths, in order to showcase its superiority over the standard CEAc in cost-effectiveness analyses with high levels of uncertainty like cancer treatments.

This simulation will utilize the hypothetical dataset generated by the authors of *Barr et al. (2023)*, and then produce both a standard CEAc and four 2CEAc's of different null hypothesis intervals. In this dataset, let $x = 0$ be the control treatment, and $x = 1$ be the new treatment. $\bar{\alpha}_{x=0} = 0.7036735$, with a 95% CI of $(0.6719546, 0.7353924)$, $\bar{\alpha}_{x=1} = 0.7266038$, with a 95% CI of $(0.6937985, 0.7594091)$, $\bar{\beta}_{x=0} = 292.1837$, with a 95% CI of $(82.77, 501.60)$, and $\bar{\beta}_{x=1} = 8272.3770$, with a 95% CI of $(5747.35, 10,797.40)$. There are 49 $x = 1$ data points and 53 $x = 0$ data points sampled. $\lambda_i = \{0, 1500000\}$ for $i \in \{1, 1001\}$. The four 2CEAc null hypothesis intervals that will be used are: $H_0 : \{-0.001 * max(\lambda), 0.001 * max(\lambda)\} = \{-1500, 1500\}$, $H_0 : \{-0.005 * max(\lambda), 0.005 * max(\lambda)\} = \{-7500, 7500\}$, $H_0 : \{-0.01 * max(\lambda), 0.01 * max(\lambda)\} = \{-15000, 15000\}$, $H_0 : \{-0.03 * max(\lambda), 0.03 * max(\lambda)\} = \{-45000, 45000\}$.

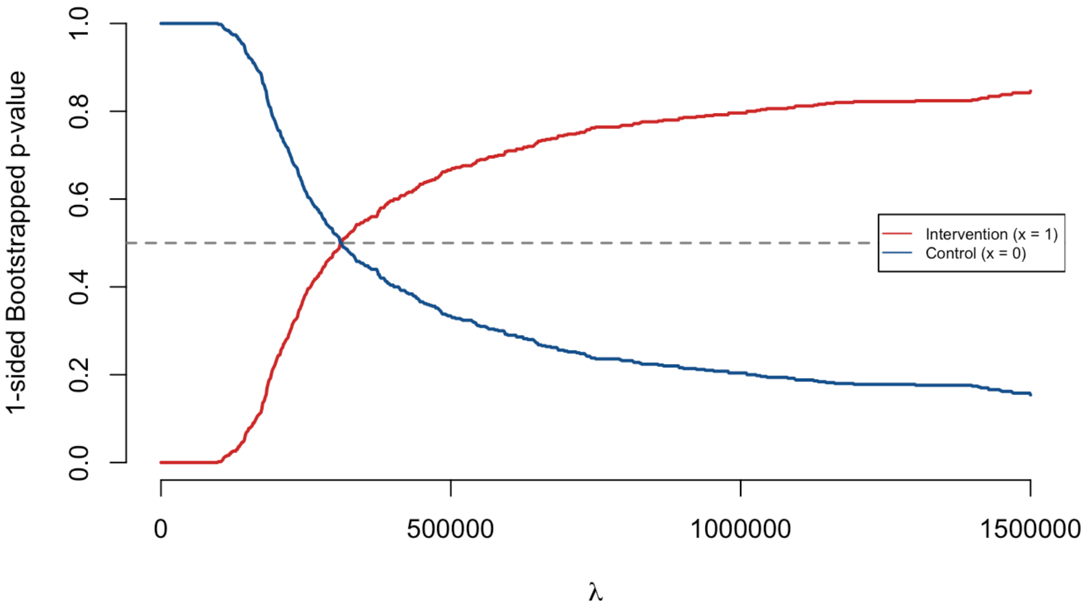

Figure 4.1: Simulated cancer treatment standard CEAc, designed to mirror the results of Barr et al.

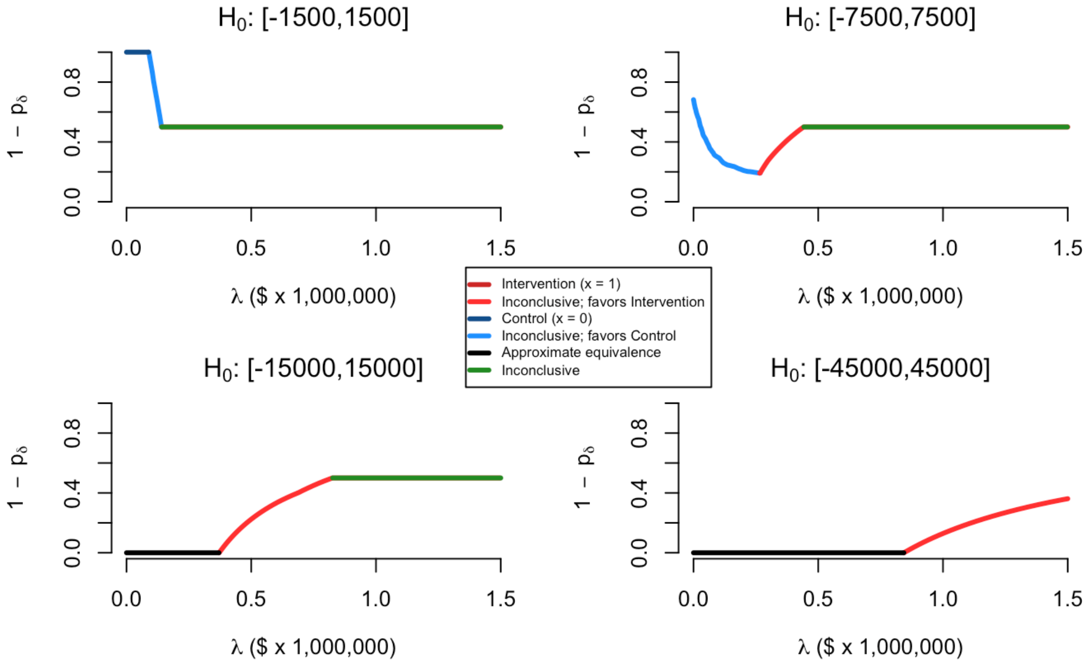

Figure 4.2: Simulated cancer treatment 2CEAc with varying null hypothesis lengths.

CHAPTER 5

Discussion

5.1 Simulation results

Figures 3.1 to 3.3 cover what happens to the 2CEAc under different simulation settings. It is important to understand the relationship between the data and 2CEAc. Figure 3.1's results for simulation A cover changes in $\alpha_{x=1}$ and $\beta_{x=1}$, where $\alpha_{x=0} = N(50,4)$, $\alpha_{x=1} = N(50+\mu,4)$, $\beta_{x=0} = N(100,160)$, $\beta_{x=1} = N(100+\gamma,160)$. Another way of viewing simulation A is differences in $\alpha_{x=1} - \alpha_{x=0}$ and $\beta_{x=1} - \beta_{x=0}$. Using the $\mu = 4$, $\gamma = 160$ 2CEAc as a reference graph, the relationship found in μ and γ changes is evident. When the difference in means between $\beta_{x=1}$ and $\beta_{x=0}$ is reduced as in the $\mu = 4, \gamma = 128$ graph, the whole curve in the 2CEAc shifts to the left. When the difference in means between $\alpha_{x=1}$ and $\alpha_{x=0}$ is reduced as in the $\mu = 3.2, \gamma = 160$ graph, the whole curve in the 2CEAc shifts to the right. Finally, when the difference in means between both the costs and the effects is reduced by the same percentage as in the $\mu = 3.2, \gamma = 128$ graph, the whole curve in the 2CEAc does not shift.

The reason for this relationship can be found in the calculations for p_δ. As explained in formula 2.1, $INMB_i = \lambda(\alpha_{x=1_i} - \alpha_{x=0i}) - (\beta_{x=1_i} - \beta_{x=0_i})$. When $\alpha_{x=1_i} - \alpha_{x=0i}$ decreases, the $INMB_i$ shrinks, which also shrinks the INMBCI used to calculate p_δ as seen in formula 2.2. Similarly, if $\alpha_{x=1_i} - \alpha_{x=0i}$ increases, the $INMB_i$ grows. When $\beta_{x=1_i} - \beta_{x=0_i}$ decreases, the $INMB_i$ grows in value, but if $\beta_{x=1_i} - \beta_{x=0_i}$ increases, $INMB_i$ decreases. Since the INMB is a ratio of differences in effects and costs, if $\alpha_{x=1_i} - \alpha_{x=0i}$ and $\beta_{x=1_i} - \beta_{x=0_i}$ decrease or increase at the same rate, then no change in $INMB_i$ is observed.

Figure 3.2's results for simulation B illustrate what happens to the 2CEAc if σ values change, or rather, what happens if the variance of the data changes. This is where $\alpha_{x=0} = N(50,\sigma)$, $\alpha_{x=1} = N(50+4,\sigma)$, $\beta_{x=0} = N(100,40\sigma)$, $\beta_{x=1} = N(100+160,40\sigma)$. As the value of σ increases, the shape of $2CEAc$ changes in a particular manner. First, the range of λ values for which $1 - p_\delta = 0$ decreases as seen from $\sigma = 4$ and $\sigma = 8$. Then, once there are no more $1 - p_\delta = 0$ values, the minimum of the graph increases up to a maximum of 0.5, as seen from $\sigma = 8$ and $\sigma = 12$. Finally, once the minimum of the 2CEAc reaches a value of 0.5, increases in σ result in an increasing range of λ values for which $1 - p_\delta = 0.5$, as seen in $\sigma = 20$.

As with many forms of statistical analysis, when the data spread increases, there tends to be more inconclusiveness. This reality applies to the 2CEAc as well. When the spread in data is small, generated confidence intervals are smaller in length, allowing for more chances for $INMBCI_j \cap H_0 = INMBCI_j$. As the spread in data increases, the generated confidence intervals increase in length, decreasing the chance for this event to occur. Once the data spread becomes too long for there to be a $INMBCI_j$ contained in H_0, the minimum for $1 - p_\delta$ increases in value. Once $1 - p_\delta$'s minimum hits 0.5, as the data spread increases, the quantity of generated confidence intervals that cover the null and alternate hypotheses equally increases in value. This makes intuitive sense, as if the data spread increases, it becomes harder to identify where interventions $x = 1$ and $x = 0$ are equivalent. With increasing data spread, we struggle to make such conclusions, as represented by the increase in λ values where $1 - p_\delta = 0.5$.

Figure 3.3's graphs for simulation C illustrate what happens to the 2CEAc when the null interval H_0 changes in length, ranging from $H_0 : \{-0.1 \times \max(\Lambda), 0.1 \times \max(\Lambda)\}$ to $\{-0.4 \times \max(\Lambda), 0.4 \times \max(\Lambda)\}$, where $\max(\Lambda) = 100$. The relationship between H_0 length and 2CEAc is simple: as the length of H_0 increases, the length of λ values for which $1 - p_\delta = 0$ increases as well.

The reasoning for this relationship is similar to that of Figure 3.2's. A smaller H_0 range means fewer opportunities for $INMBCI_j \cap H_0 = INMBCI_j$. This is why it is important to choose a reasonable null interval range (like H_0 : $\{-0.15 \times \max(\Lambda), 0.15 \times \max(\Lambda)\}$) in setting up the 2CEAc. If the interval length is too small, one would miss out on λ values for which the interventions are nearly equivalent cost-effectiveness wise. If the interval length is too large, then the 2CEAc constructed would only identify immense cost-effectiveness differences in $x = 1$ and $x = 0$. Ultimately, null interval choice determines how much of a cost-effectiveness difference in $x = 1$ and $x = 0$ is necessary to be considered significant. This is why it is very important to specify the null hypothesis interval (and conversely, the level of the confidence interval) before creating the 2CEAc. Not doing so runs the risk of the researcher shopping for intervals to get a better conclusion.

Figures 3.4-3.6 cover differences between the standard CEAc (labeled as First Generation) and the 2CEAc (labeled as Second Generation) in three separate circumstances. All three figures follow $\alpha_{x=0} = N(50, \sigma)$, $\alpha_{x=1} = N(50+4, \sigma)$, $\beta_{x=0} = N(100, 40\sigma)$, $\beta_{x=1} = N(100+160, 40\sigma)$. Figure 3.4 covers an example using a standard dataset that carries no significant quirks (where $\sigma = 4$). The range of λ values for which $1 - p_\delta = 0$ in the 2CEAc are equivalent to the range of λ values for which the standard CEAc doesn't have an identified 100% optimal intervention. λ values for which $1 - p_\delta = 1$ in the 2CEAc line up with values for which the 1-sided bootstrapped p-value is 1 in the standard CEAc.

With a null interval of $H_0 : \{-0.15 \times \max(\Lambda), 0.15 \times \max(\Lambda)\} = \{-15, 15\}$, Figure 3.4 is what we would expect as result for a 2CEAc. The 2CEAc captures all of the λ values for which we could consider the interventions equivalent for the standard CEAc. This is one of the advantages of using a 2CEAc over the standard CEAc. The standard CEAc does answer when one intervention appears to be optimal, but it struggles to identify when both interventions are equivalent in terms of cost-effectiveness. The 2CEAc does not have this problem, clearly identifying the range of λ values for which it believes that both interventions are equivalent in terms of cost-effectiveness.

Figures 3.5 and 3.6 cover simulations E and F respectively, which entail using datasets with quirks. Simulation E covers a scenario where $\sigma = 20$, or rather, a scenario where the data spread is high. Simulation F covers a scenario where instead of each intervention having a sample size of 10000, as seen in the rest of the simulations, each intervention instead has a sample size of $\frac{N}{2} = 400$. Both simulations have similar results. The standard CEAc for Figures 3.5

and 3.6 have a smaller slope as compared to Figure 3.4's. The 2CEAc for Figures 3.5 and 3.6 do not have any $1 - p_\delta$ values smaller than 0.5, and the λ values for which $1 - p_\delta = 0.5$ cover the majority of λ values for which the standard CEAc does not have a intervention with bootstrapped p-value nearing 1. Finally, it appears that Figure 3.5's standard CEAc shifted slightly to the right.

The standard CEAc's shift to the right in Figure 3.5 is explained by the high data spread. It should be noted that rather than claiming that there is an equivalent chance between interventions $x = 1$ and $x = 0$ being optimal at $\lambda = 45$ as seen in the standard CEAc, the 2CEAc states that the results are inconclusive. Both simulations E and F have a similar problem: the data given forces a wider range of λ values for which either $x = 1$ or $x = 0$ could potentially be the optimal intervention. The standard CEAc handles this problem by decreasing the size of the slope, which increases the number of λ values for which the choice of intervention necessitates consideration (that is, λ values for which the bootstrapped p-value is not 0). The 2CEAc handles this problem differently, instead claiming a range of λ values for which no significant conclusions can be made. This is one of the advantages the 2CEAc has over the standard CEAc: if the data cannot support a significant conclusion at a λ value, rather than giving a pair of close probabilities like in the standard CEAc, the 2CEAc outright states the reality at hand.

5.2 Applied cancer results

As explained in Section 2.3, cancer data sets and their analyses can be limited by the lack of a significant quantity of data, whether it be due to high costs, or the rarity of the specific health scenarios being focused on. Further, the differences in effects can be minor. These scenarios leave statistical analyses with a high amount of uncertainty, which as discussed in *Barr et al. (2023)*, are something that is challenging to depict in a digestible manner. This is where the 2CEAc best excels in cost-effectiveness as compared to the standard CEAc, by efficitively contextualizing uncertainty [8].

Figure 3.7 is very similar to the standard CEAc generated in *Barr et al. (2023)*, showing a scenario where there is no 100% certainty of one treatment being cost-effectively optimal when $\lambda > 0$. While the standard CEAc does incorporate uncertainty into having p-values less than 1, it leaves the reader lacking the comprehension necessary to make concrete decisions, leading to the following example questions:

- How high (or low) must the p-values be in order to make meaningful conclusions?
- Would it be right for a statistician to suggest the new treatment at $\lambda = 100000$, where the p-value is about 0.9?
- Further, for what λ values are the treatments equivalent?

Certainly there is an intersection point at a single λ value, but beyond that, the answer is unclear. Finally, as said in *Barr et al. (2023)*, "the CEAC does not quantify how cost effective a new treatment is, it shows only the probability that a new treatment is cost-effective" [8].

The 2CEAc addresses all of the problems posed above. The 2CEAc allows the statistician to quantify a minimum for how better a medical intervention must be than its competitor in cost-effectiveness through the employment of null hypothesis intervals. As seen in Figure 3.8, changing the width of the null hypothesis interval greatly changes what is depicted graphically. By widening the null interval, an increasing number of λ values have both treatments equivalent in cost-effectiveness (as shown in Figure 3.3, and discussed Section 3.1). This addresses the standard CEAc's flaw in lacking a clarified range of values for which the treatments are equivalent. Perhaps most importantly for the purposes of this simulation, the 2CEAc effectively depicts uncertainty and inconclusiveness in an understandable manner. The horizontal green lines marking pure inconclusiveness give the reader a clear indication that no claim could be made on the relative cost-effectiveness of the treatments given the data.

Ultimately, the purpose of both the 2CEAc and standard CEAc is to help the statistician determine if a medical intervention is optimal in cost-effectiveness as compared to a control intervention over a range of λ values. When the data clearly supports one intervention over another, the graphs are similar in effectiveness. When high degrees of uncertainty are involved as in this cancer example however, the 2CEAc is superior in depicting equivalence and inconclusiveness.

5.3 The case for a second generation cost-effectiveness acceptability curve

The main purpose of using a standard CEAc is to identify which intervention is optimal in cost-effectiveness over a range of possible λ values, which it does by directly calculating standard p-values. While this graph is effective when looking at λ values for which the probability is either near 1 or 0, when the p-values of both interventions are closer in value, interpretations become muddled. If intervention $x = 1$ has a p-value of 0.85, and $x = 0$ a p-value of 0.15 for example, the question remains as to whether the practitioner could reasonably suggest that intervention $x = 1$ should be chosen, or if they should say the result is simply inconclusive. More critically, the standard CEAc mostly fails to identify points where the two interventions are near-equivalent in terms of cost-effectiveness. The intersection point only returns a single lambda value, which is insufficient.

The 2CEAc does everything the standard CEAc excels at, while rectifying the issues previously discussed. The 2CEAc identifies λ values where one can conclude there is evidence for one intervention being optimal in cost-effectiveness. If the data is inconclusive at specific λ values, the 2CEAc clearly communicates this reality (as seen in Figures 3.5 and 3.6). Finally, and most importantly, the 2CEAc identifies λ values for which the interventions are considered near-equivalent in cost-effectiveness (as seen in Figure 3.4). Further, one can also actively change what "near-equivalent" means, according to their own needs (as seen in Figure 3.3). Such changes should only occur before beginning the creation of the 2CEAc.

All of these advantages the 2CEAc has over the standard CEAc descend from the benefits of using a second generation p-value over a standard p-value. The use of null hypothesis intervals enables a more robust analysis that enables results clearer than simply using two standard p-values (which is what the standard CEAc does). Like the standard CEAc, understanding the 2CEAc is simple, as shown in Section 2.1. Ultimately, when conducted effectively, the 2CEAc is a superior form of cost-effectiveness analysis as compared to the standard CEAc.

CHAPTER 6

Appendix

6.1 Second generation acceptability curve psuedocode

The procedure for creating a second generation acceptability curve consists of calculating second generation p-values, and graphing by comparing it to the lambda values supplied. The steps described will mirror Chapter 2's writings. It is assumed that the practitioner would have a data set consisting of sampled costs and benefits for two distinct interventions.

First, after loading a seed, one would define multiple values based on the numerical scale of the data: the length of the null hypothesis centered around zero (`nullin`), the number of bootstraps (`B`), and the range and quantity of λ values (`ls`). With these values determined, the first step is to bootstrap and calculate INMBs. In this paper's case, a simple replacement bootstrap resamples the data points and applied formula 2.1: `INMBs = ls * (difference in mean effects) - (difference in mean costs)`. This produces one INMB for each unique λ value, which is stored in `INMBs`. This process is repeated `B` times.

A 95% confidence interval is then calculated for the bootstrapped INMBs by finding the 2.5% and 97.5% `quantile` for each lambda value. The lengths of the null hypothesis (`length.null`), range of λ values (`length.inters`), and confidence intervals (`length.CI`) are then calculated. For each INMB confidence interval for a λ value in `ls`, the second generation p-value is calculated by formula 2.2: `pdeltas = (length.inters/length.CI) * pmax(length.CI/(2 * length.null), 1)`. All needed variables are now calculated, and all that remains is depicting the curve.

Following the process described in 2.1.2, each λ's INMB confidence interval is compared to the null hypothesis. Using a categorical variable `whichtrt`, each λ is assigned a group. This assignment process can be accomplished in a variety of ways. This paper first assigned each λ value's `whichtrt` to an intervention, based on whether the confidence interval ranges more below or above the null hypothesis interval. Then, the second generation p-values `pdeltas` is taken into account as follows: if `1- pdeltas = 0`, then that λ's `whichtrt` is set to approximate equivalence; if `1- pdeltas = 0.5`, then that λ's `whichtrt` is set to inconclusiveness; and if `1> 1- pdeltas > 0`, then that λ's `whichtrt` is set to inconclusiveness favoring the previously assigned intervention.

The final step is graphing the second generation analysis curve, by plotting `ls` vs `1 - pdeltas`, with the line being colored according to `whichtrt`, and a legend being supplied for comprehension.

References

[1] Fenwick E, Marshall DA, Levy AR, Nichol G. Using and interpreting cost-effectiveness acceptability curves: an example using data from a trial of management strategies for atrial fibrillation. BMC Health Serv Res. 2006;6:52. Published 2006 Apr 19.

[2] Bang H, Zhao H. Median-Based Incremental Cost-Effectiveness Ratio (ICER). J Stat Theory Pract. 2012;6(3):428-442.

[3] Antonides CFJ, Cohen DJ, Osnabrugge RLJ. Statistical primer: a cost-effectiveness analysis. Eur J Cardiothorac Surg. 2018;54(2):209-213.

[4] Fenwick E, Claxton K, Sculpher M. Representing uncertainty: the role of cost-effectiveness acceptability curves. Health Econ. 2001;10(8):779-787.

[5] Fenwick, E. and O'Brien, B.J. and Briggs, A. Cost-effectiveness acceptability curves - facts, fallacies and frequently asked questions. Health Econ. 2001; 13(5):405-415.

[6] Blume JD, Greevy RA Jr, Welty VF, Smith JR, Dupont WD. An Introduction to Second-Generation p-Values. The Amer Stat. 2019;73:sup1, 157-167

[7] Blume JD, D'Agostino McGowan L, Dupont WD, Greevy RA Jr. Second-generation p-values: Improved rigor, reproducibility, & transparency in statistical analyses. PLoS One. 2018;13(3):e0188299. Published 2018 Mar 22.

[8] Barr HK, Guggenbickler AM, Hoch JS, Dewa CS. Real-World Cost-Effectiveness Analysis: How Much Uncertainty Is in the Results?. Curr Oncol. 2023;30(4):4078-4093. Published 2023 Apr 7.

[9] Sendi P. Dealing with Bad Risk in Cost-Effectiveness Analysis: The Cost-Effectiveness Risk-Aversion Curve. Pharmacoeconomics. 2021;39(2):161-169.

[10] Zheng P, Liu J. Cost-Effectiveness Analysis of Hp and New Gastric Cancer Screening Scoring System for Screening and Prevention of Gastric Cancer. Curr Oncol. 2023;30(1):1132-1145. Published 2023 Jan 13.

[11] Ferket BS, Oxman JM, Iribarne A, Gelijns AC, Moskowitz AJ. Cost-effectiveness analysis in cardiac surgery: A review of its concepts and methodologies. J Thorac Cardiovasc Surg. 2018;155(4):1671-1681.e11.

[12] Ferrari G, Torres-Rueda S, Chirwa E, et al. Prevention of violence against women and girls: A cost-effectiveness study across 6 low- and middle-income countries. PLoS Med. 2022;19(3):e1003827. Published 2022 Mar 24.

[13] Selva-Sevilla C, Fernández-Ginés FD, Cortiñas-Sáenz M, Gerónimo-Pardo M. Cost-effectiveness analysis of domiciliary topical sevoflurane for painful leg ulcers. PLoS One. 2021;16(9):e0257494. Published 2021 Sep 20.

[14] Briggs AH, O'Brien BJ, Blackhouse G. Thinking outside the box: recent advances in the analysis and presentation of uncertainty in cost-effectiveness studies. Annu Rev Public Health. 2002;23:377-401.

[15] Herzel BJ, Samuel SP, Bulfone TC, Raj CS, Lewin M, Kahn JG. Snakebite: An Exploratory Cost-Effectiveness Analysis of Adjunct Treatment Strategies. Am J Trop Med Hyg. 2018;99(2):404-412.

[16] Adashek JJ, Genovese G, Tannir NM, Msaouel P. Recent advancements in the treatment of metastatic clear cell renal cell carcinoma: A review of the evidence using second-generation p-values. Cancer Treat Res Commun. Published online January 3, 2020.

[17] Stewart TG, Blume JD. Second-Generation P-Values, Shrinkage, and Regularized Models. Front. Ecol. Evol. 2019;7:486.