# Visualizing Cost Effectiveness Analysis with Second-Generation Acceptability Curves

Ashley E. Mullan[1], Nicholas G. Micheletti[1], Jeffrey Blume[3], Lauren R. Samuels[2], Andrew J. Spieker[2]
1. Vanderbilt University – Department of Biostatistics, Nashville, TN
2. Vanderbilt University Medical Center – Department of Biostatistics, Nashville, TN
3. University of Virginia – School of Data Science, Charlottesville, VA

## INTRODUCTION

- Cost effectiveness analysis aims to find monetarily efficient medical interventions, and efficiency is often visualized with **acceptability curves**.
- These curves depict some function of a given **effectiveness metric** for various budget constraints.
- Common metrics for comparing two treatments include the incremental cost-effectiveness ratio and the **net monetary benefit**, which considers the finite nature of the budget.
- Previous works use a **bootstrapped p-value** as the function of the effectiveness metric, which often confounds effect size and precision[1].
- The **second-generation p-value**[2] replaces the standard point null hypothesis with an **interval null hypothesis** to allow a more intuitive interpretation, the **proportion of hypotheses** belonging to the null interval.
- Values close to 0 or 1 imply a preference for one treatment over another, and the degree of **inconclusiveness** increases as values approach 0.5.
- **Second-generation acceptability curves** use this method to draw conclusions about treatment optimality rather than the traditional p-value.

## METHODS

We demonstrate properties of the second-generation acceptability curve using simulated data. Specifically, we explore how **different sample sizes** and **null interval lengths** affect the **shape** of the second-generation curve and the practical **conclusions** the curves imply.

The effectiveness metric under consideration is the **incremental net monetary benefit (INMB)**, computed for two competing treatments $A_0$ and $A_1$. We represent the limit of available resources as $\lambda$, the mean of treatment effect $Z$ for group $A_i$ as $\alpha_i$, and the mean of cost $Y$ for group $A_i$ as $\beta_i$.

$$\lambda \times (\alpha_1 - \alpha_0) - (\beta_1 - \beta_0)$$

The second-generation p-value $p_\delta$ is computed using Blume's formula[2] (2018), where $H_0$ represents the null interval and $I_C$ the confidence interval for the INMB.

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

Additionally, we provide a freely available **R package** to compute second-generation p-values for a given dataset and create second-generation acceptability curves. The curves returned by the package are **ggplot2** objects and are therefore able to be further customized for a user's needs.

## RESULTS

### Package Tutorial

A basic use case of the package only requires four function calls! The user prepares the data, constructs confidence intervals for the incremental net monetary benefit, uses those results to generate second-generation p-values, and then creates the plot.

```
#install and load packages
library(devtools)
install_github(repo = "ashleymullan/visegen")
library(visegen)

#working example
df <- prep_data(A, Y, Z)
cis <- ci_nmb(df, lambdas)
pvals <- compute_2gp(null_scale, lambdas,
                     cis$lower_bounds, cis$upper_bounds)
plot_2gp(lambdas, null_scale,
         cis$lower_bounds, cis$upper_bounds,
         p_vals)
```

### Simulation Setup

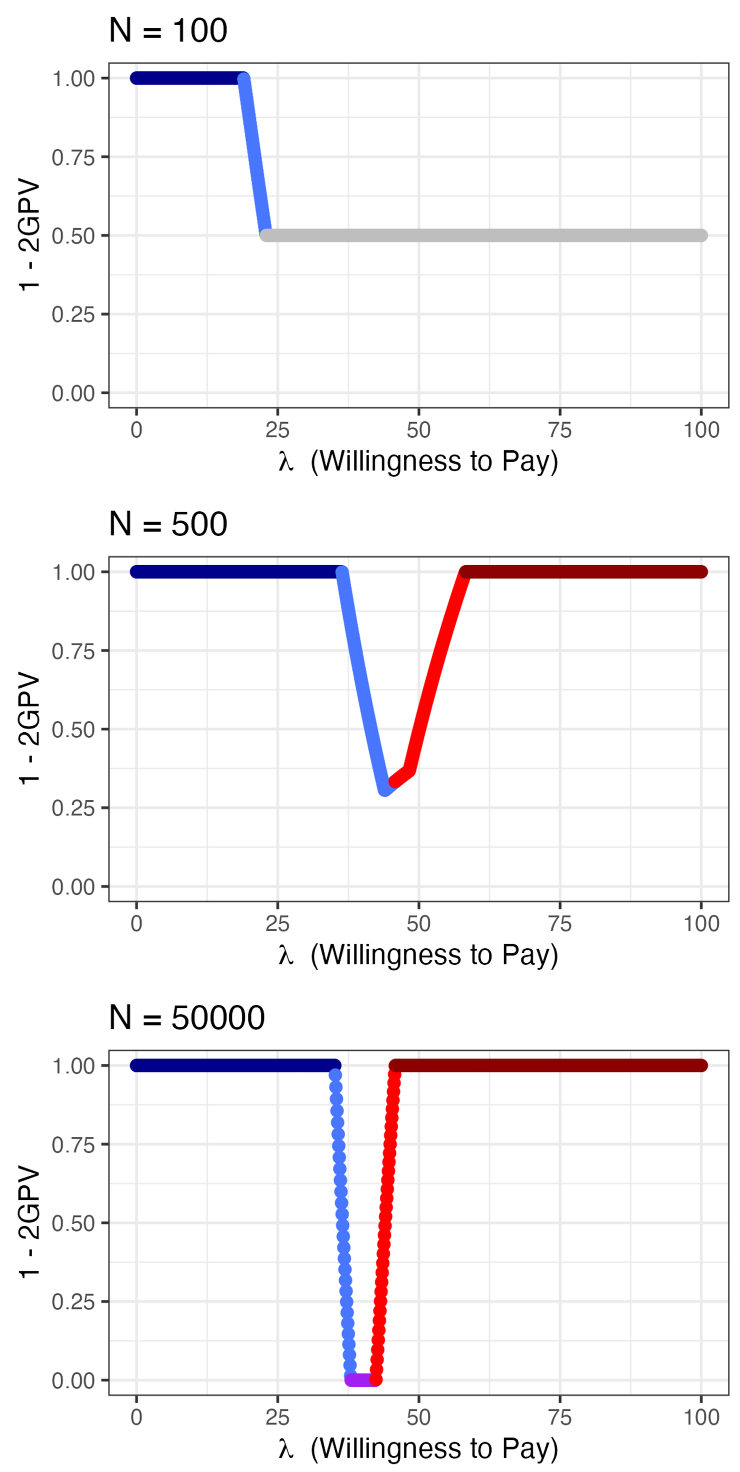
For both studies, we specify the following:

- Benefit: $Z \sim \text{Normal}(50 + 4 \times \text{Treatment}, 9^2)$
- Cost: $Y \sim \text{Normal}(100 + 160 \times \text{Treatment}, 10^2)$
- Willingness to Pay: $\lambda \in [0,100]$

For Study 1, we **fix** the null interval to have width 90 and **vary** the sample size to consider N ∈ {1000, 5000, 25000}.

For Study 2, we **fix** the sample size N = 5000 and **vary** the null interval to be be scaled by factors of {0.05, 0.45, 0.75}, yielding null interval widths of {10, 90, 150}.

### Varied Sample Size

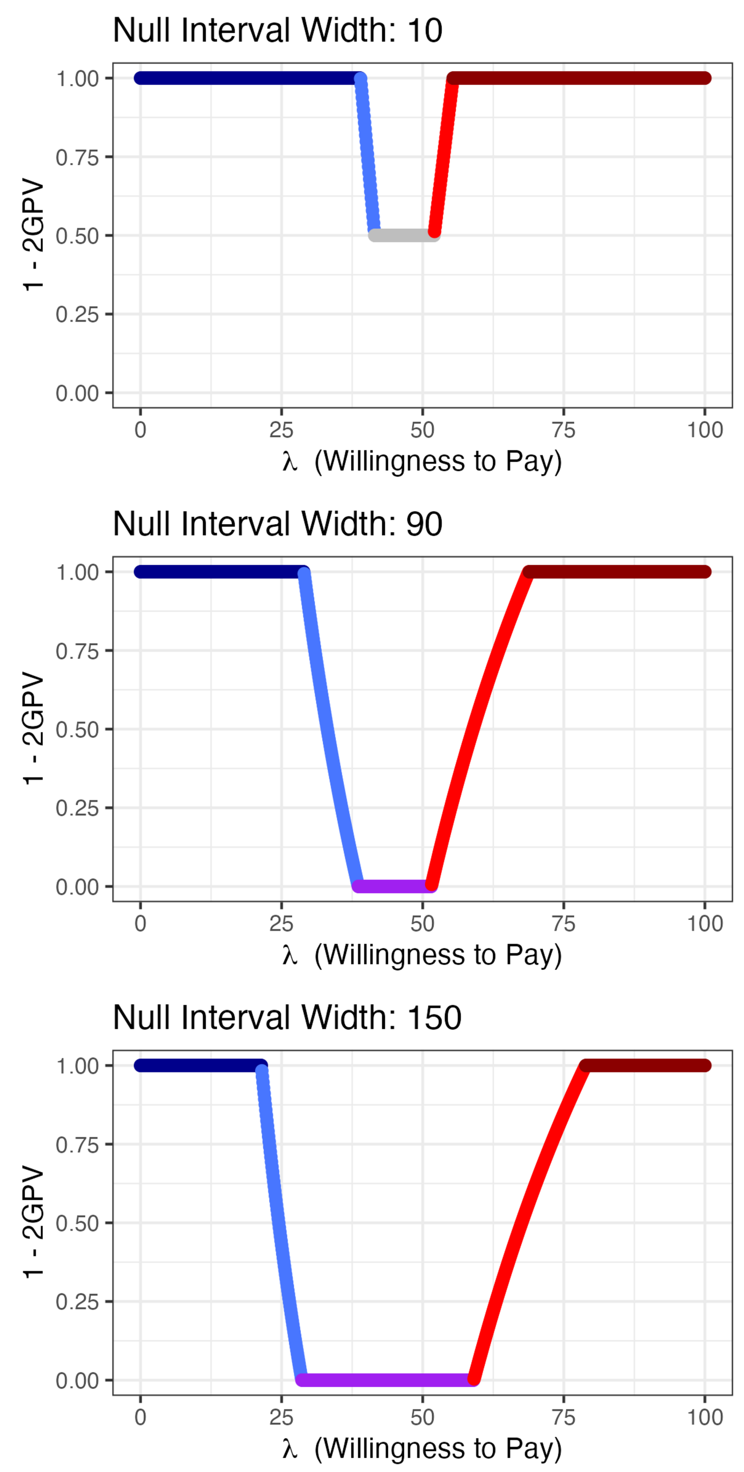



At our smallest sample size, we see mostly inconclusive results and never see a preference for the treatment under the given willingness to pay.

As sample size grows, we start to gain more intuition as the region of inconclusiveness disappears.

If sample size grows large enough, we eventually start to see a small region of equivalence between treatment groups.

### Varied Null Interval




With narrower null interval widths, the p-values stay firmly in the bottom half of the domain, and we do not see a region of equivalence.

Increasing the width of the null interval allows the p-values to span the full possible range and we begin to see a region of equivalence.

As the width of the null interval expands, so does the region of equivalence between treatments.

## DISCUSSION AND CONCLUSION

- The clear depiction of **inconclusive and equivalent results** that a practitioner can achieve with second-generation curves **is not possible** using **first-generation** curves, so the **second-generation** version may be advantageous especially in **the pilot study phase**, where not much is known about treatment benefits.
- The **shape** of a second-generation accessibility curve **varies** noticeably over the range of possible budget constraints with **sample size** and with **the width of the null interval**.
- Only one range of **willingness to pay** (limit of available resources) was considered, but some scenarios may have different curve behavior if the **range is extended**, especially in situations with lower sample size.
- As **sample size grows**, we eventually gain the precision to **cover all possible** second-generation p-values.
- As the **null interval gets wider**, we again start to cover all possible second-generation p-values (demonstrating regions of optimality, inconclusiveness, and equivalence) but begin to see more prominent **regions of equivalence** between treatments.
- Limitations of the second-generation acceptability curve include a lack of a simple way to depict **the numerical degree of inconclusiveness**.
- Future functionality for the `visegen` package may include options to **visually compare** first and second-generation acceptability curve conclusions and adding other **computational options** to consider **treatment optimality**, such as the overlapping method.

## REFERENCES

1. Blume, J. D., Greevy, R. A., Welty, V. F., Smith, J. R., & Dupont, W. D. (2019). An Introduction to Second-Generation *p*-Values. *The American Statistician*, *73*(sup1), 157–167. https://doi.org/10.1080/00031305.2018.1537893
2. Blume, J. D., D'Agostino McGowan, L., Dupont, W. D., & Greevy, R. A. (2018). Second-generation P-values: Improved rigor, reproducibility, & transparency in statistical analyses. *PLOS ONE*, *13*(3). https://doi.org/10.1371/journal.pone.0188299

Scan for poster!

Scan for package!