



# Relative Efficacy of Virtual and In-Office Conservative Care for Musculoskeletal Conditions

A Propensity Score Matching-Based Analysis

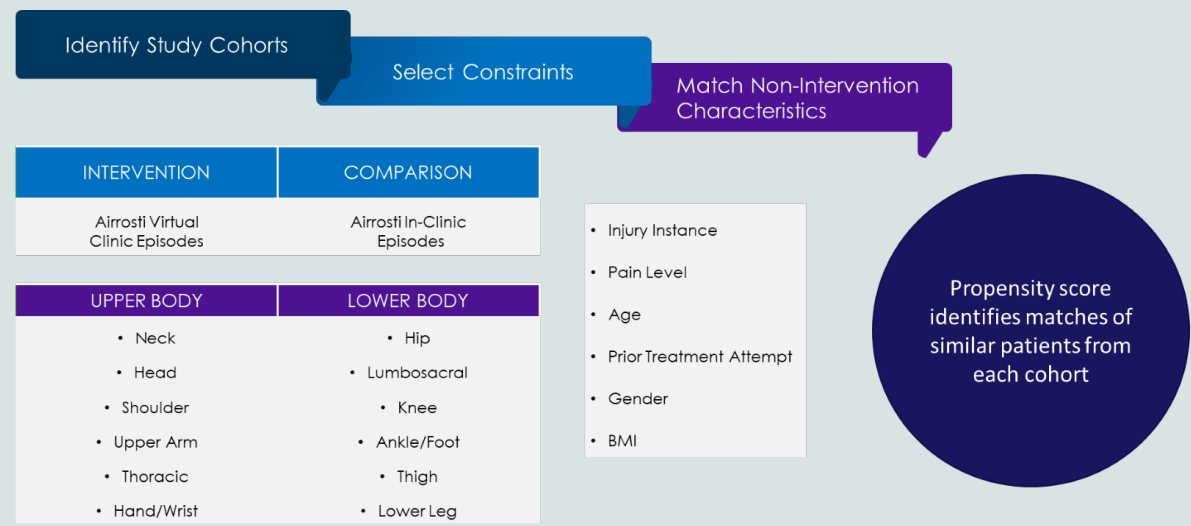
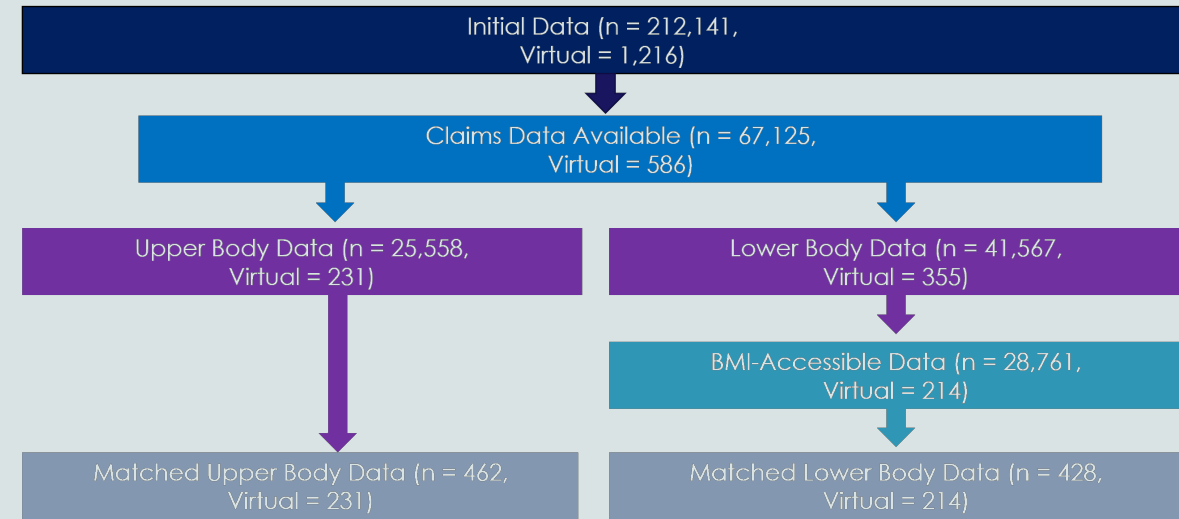
# Comparative Evaluation of Virtual Health Options

As more and more patients seek out virtual care and an increasingly large number of providers seek to accommodate this demand, it is necessary to evaluate its effectiveness relative to traditional, in-office care. However, data with which to carry out this evaluation is scarce due to the short time frame during which virtual care options have been adopted. It is therefore necessary to locate and apply statistical techniques with which to carry out this analysis despite the limited nature of the data.

This white paper presents a technique for using propensity score matching to carry out this evaluation alongside the results of this analysis among Airrosti patients seeking conservative Musculoskeletal (MSK) care, which demonstrated that **virtual care patients achieve similar positive outcomes to in-office patients.**

# Overview of Study Design

- Limit data to cases where key variables are available
- Control for confounding variables using propensity score matching, limited to same area of injury
- Evaluate the effectiveness of Airrosti Remote Recovery, Airrosti’s virtual conservative MSK care model (physiotherapy), for upper and lower body injuries on pain improvement, surgical avoidance, injury resolution and visit completion.



# Propensity Score Matching

- Different groups of people may choose virtual conservative MSK care than in-office. This presents a potential confounder to analysis of the relative effectiveness of the two treatment options.
- Propensity score matching addresses these unknown confounders by mimicking the effects of a randomized control trial, matching patients with patients who are similar in variables deemed likely to affect outcomes. <sup>1</sup>

# Building a Propensity Score Model

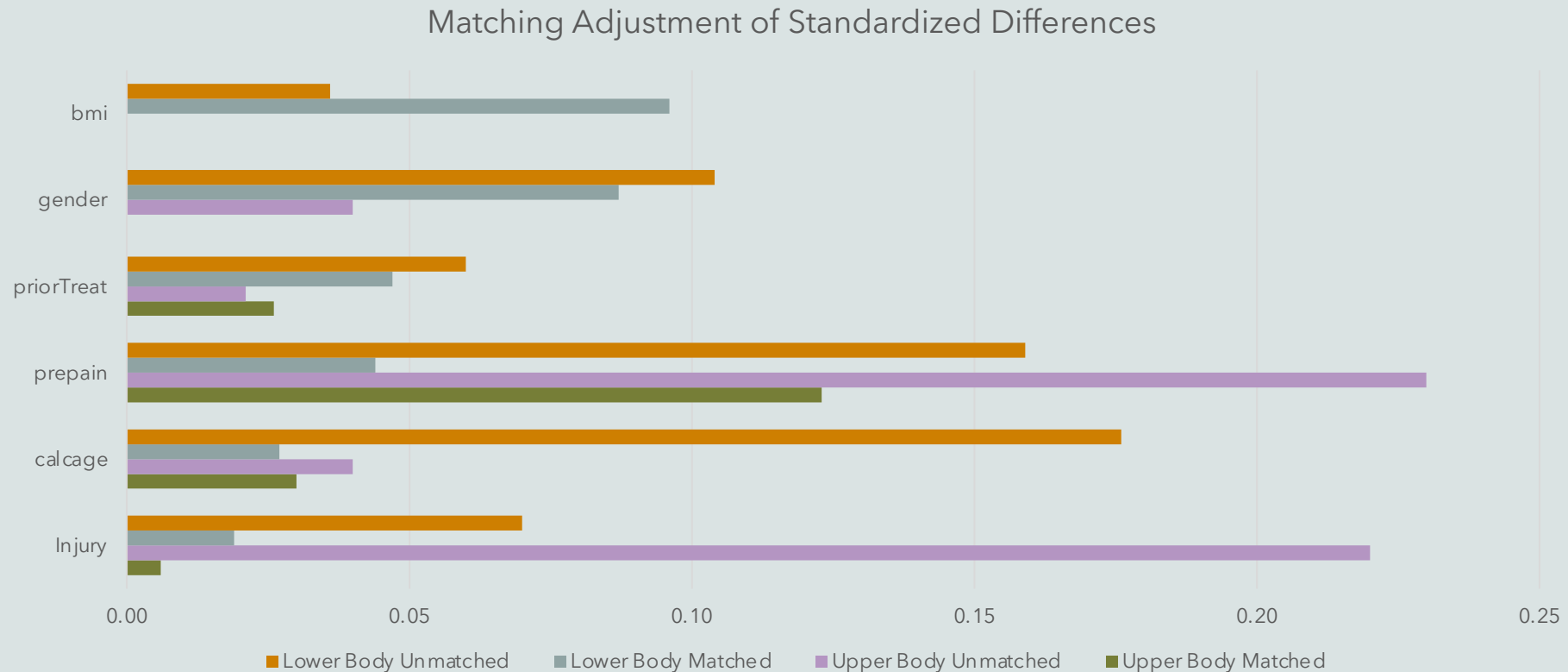
- We matched subjects on: Injury location, instance of injury, initial reported pain level (1-5), age, presence or absence of prior treatment attempts, gender, and BMI (for lower-body, weight-bearing injuries only). Matching was accomplished via the optmatch R package implementing the RELAX-IV algorithm.<sup>2</sup> These values were selected based on available data and characteristics controlled for in prior studies of musculoskeletal injury.<sup>3</sup>

# Categorizing Injury Location

- Due to a preponderance of evidence that injuries in different locations have different outcomes, virtual patients were only permitted to match with in-office patients whose injury was in the same location.
- Upper body and lower body (I.E, weight-bearing) injuries were analyzed separately, due to the highly disparate effects of BMI on outcomes for those two groups.
- Within these groups, injuries were categorized into Hip, Lumbar/Sacral, Knee, Ankle/Foot, Thigh, Lower Leg (Lower Body) and Neck, Head, Shoulder, Upper Arm, Thoracic, Elbow, Hand/Wrist (Upper Body).

# Propensity Score Evaluation

- Mean standardized differences of model variables were used to evaluate the success of matching. For almost all variables, the matching process lowered the standardized difference below 0.1. Values above 0.1 are considered potential problems, while values below it mean analysis can be confidently continued.<sup>4</sup>



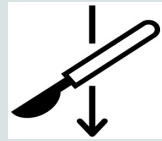
# In-Clinic Claim Study Results

Independent third-party claims studies from Milliman MedInsight and Koan Health present the high-level outcomes of Airrosti in-clinic patients. These studies reviewed up to a billion claims and over 2 million episodes of care using Blue Health Intelligence Data.<sup>5</sup>



**43%**

Reduction in Total  
Cost of Care



**80%**

Reduction in  
Surgical Utilization



**67%**

Reduction in High-Tech  
Imaging Utilization



**55%**

Reduction in  
Episode Length

## The Airrosti Difference

**1M+**

Patient Cases

**99.6%**

Patient Satisfaction

**88%**

Injury Resolution

**91%**

Prevented Further Medical Services



# Evaluating The Matched Dataset

- Single parametric tests (t-tests) are typically used to compare the degree and significance of differences between two datasets.<sup>6</sup> For our purposes, however, we want to examine the level of *equivalence*. We therefore applied a Two One-sided T Test (TOST) to the match data, which determines the maximum possible degree and direction of the difference in outcome between two conditions.<sup>6</sup>



# VARIABLE DEFINITIONS

- Patient outcomes were evaluated using variables selected from prior reviews of the subject.<sup>8</sup> These were:
  - **Pain Improvement:** The difference between initial and final reported pain on a 5- point scale
  - **Visit Completion:** The absolute number of Airrosti visits completed
  - **Surgery Avoidance:** Whether the patient reported avoiding a considered or scheduled surgery based on their Airrosti results
  - **Injury Fixed:** Whether the patient reported their injury as 'fixed' in post-therapy surveys

# Overview of Results

Without having access to all conceivable data, it is impossible to prove that the effect of two conditions are *exactly* equivalent. We therefore defined "equivalence" in this case as 95% confidence that the difference between the mean result for the virtual and in-office conditions was less than one half of one standard deviation.

At this level of difference, there is little meaningful effect on individuals.<sup>9</sup>

<b>Variable</b>	<b>Upper Body Result</b>	<b>Lower Body Result</b>
Pain Improvement	Virtual is Equivalent or Superior	Virtual is Equivalent or Inferior
Visit Completion	Virtual is Equivalent or Superior	Virtual is Equivalent or Superior
Surgery Avoidance	Virtual is Equivalent or Superior	Virtual is Equivalent or Superior
Injury Fixed	Virtual is Equivalent	Virtual is Equivalent

# Understanding TOST Plots



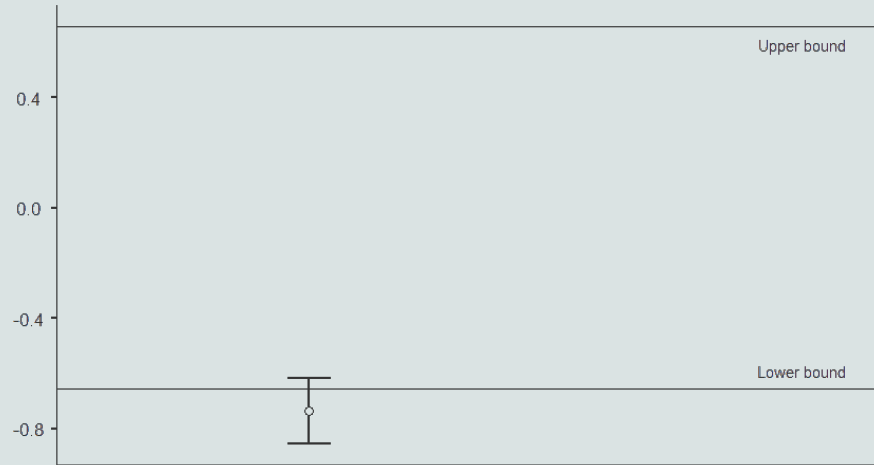
If there is no noteworthy ( $> \frac{1}{2}$  a standard deviation) difference between virtual and in-office patients, the confidence interval will fall in this range. Confidence intervals above this range would indicate that virtual patients performed worse.

The confidence interval for the difference between virtual and in-office patients. Negative numbers represent better performance for virtual.

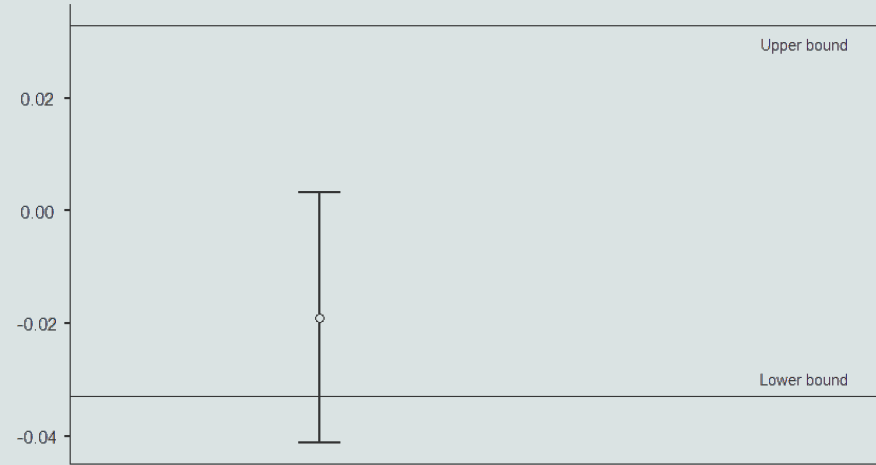
Because the range falls partially between the upper and lower bound, and partially below the lower bound, we can say that virtual patients performed equal to or better than in-office patients with 95% confidence.

# TOST Results: Upper Body

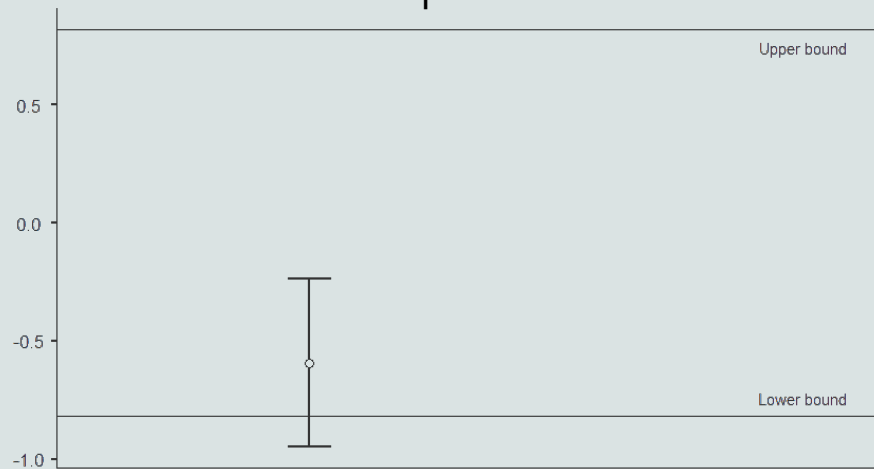
## Pain Improvement



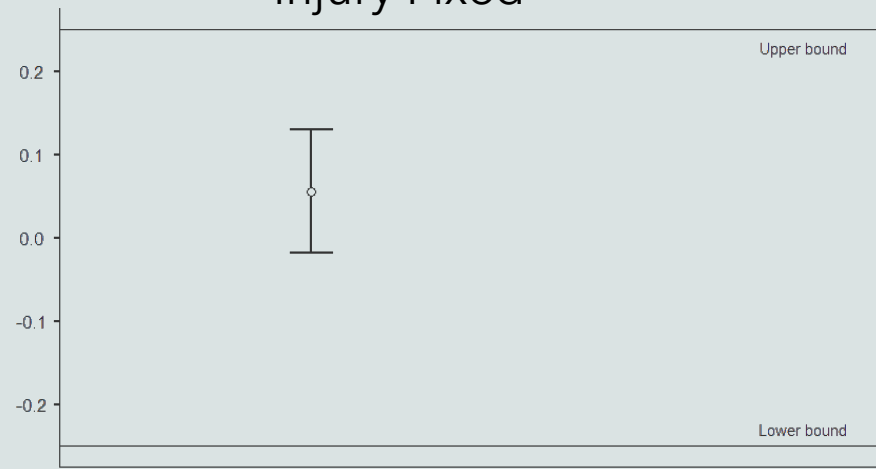
## Surgery Avoidance



## Visit Completion

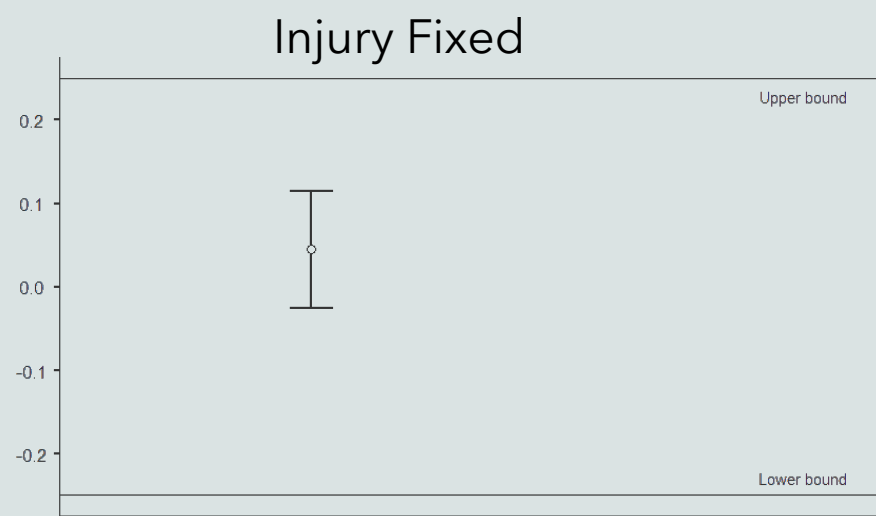
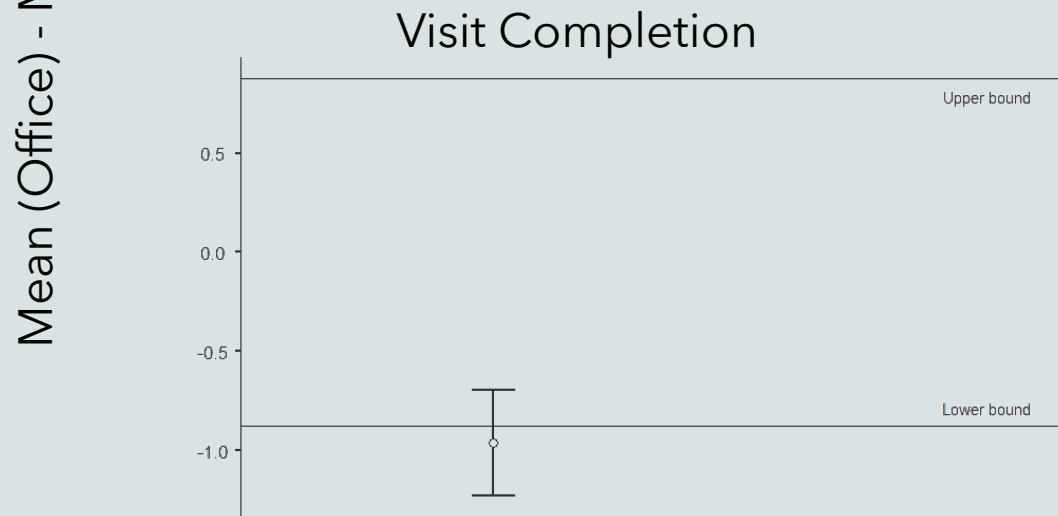
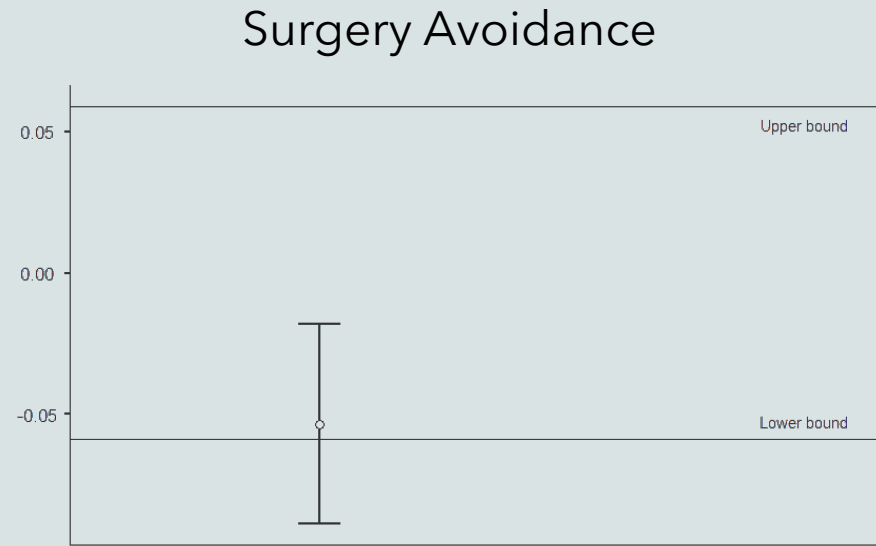
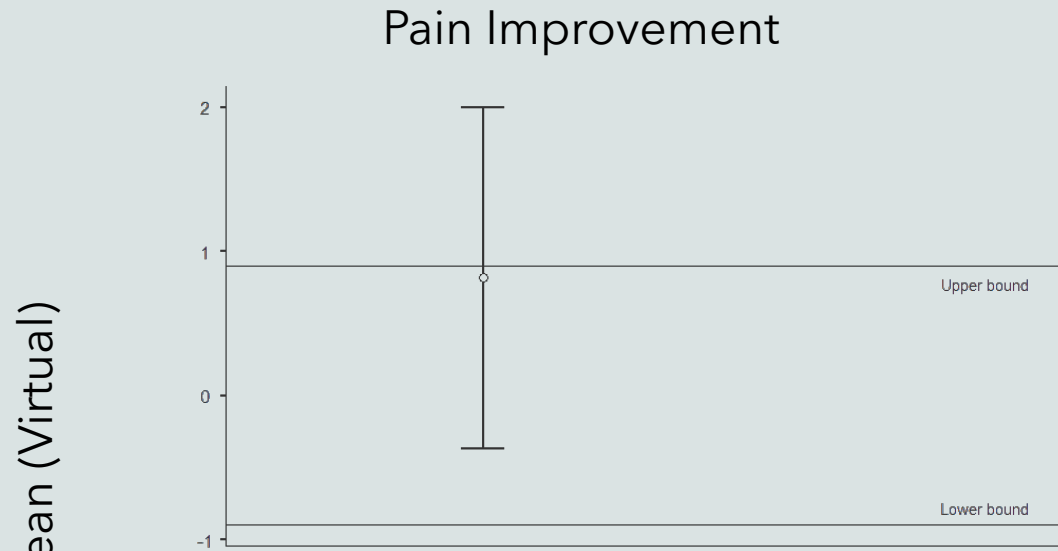


## Injury Fixed



Mean (Office) - Mean (Virtual)

# TOST Results: Lower Body



# Bibliography

- 1: Williamson EJ, Forbes A. Introduction to propensity scores. *Respirology*. 2014 Jul;19(5):625-35. doi: 10.1111/resp.12312. Epub 2014 May 29. PMID: 24889820.
- 2: Bertsekas, D., and Tseng, P. ~1994!. "RELAX-IV: A faster version of the RELAX code for solving minimum cost flow problems." Completion Report under NSF Grant CCR-9103804, Department of Electrical Engineering and Computer Science, MIT, Cambridge, Mass
- 3: Peat G, McCarney R, Croft P. Knee pain and osteoarthritis in older adults: a review of community burden and current use of primary healthcare. *Ann Rheum Dis* 2001 Feb; 60(2):91-97
- 4: Andrade C. Mean Difference, Standardized Mean Difference (SMD), and Their Use in Meta-Analysis: As Simple as It Gets. *J Clin Psychiatry*. 2020 Sep 22;81(5):20f13681. doi: 10.4088/JCP.20f13681. PMID: 32965803.
- 5: Study completed by Koan Health utilizing Milliman MedInsight developed methodology and BHI claims from BCBSTX paid between Feb. 2012 and Dec. 2020.
- 6: Jankowski KRB, Flannelly KJ, Flannelly LT. The t-test: An Influential Inferential Tool in Chaplaincy and Other Healthcare Research. *J Health Care Chaplain*. 2018 Jan-Mar;24(1):30-39. doi: 10.1080/08854726.2017.1335050. Epub 2017 Jun 16. PMID: 28622103.
- 7: Shieh G. Exact Power and Sample Size Calculations for the Two One-Sided Tests of Equivalence. *PLoS One*. 2016 Sep 6;11(9):e0162093. doi: 10.1371 PMID: 27598468
- 8: Peat G, McCarney R, Croft P. Knee pain and osteoarthritis in older adults: a review of community burden and current use of primary healthcare. *Ann Rheum Dis* 2001 Feb; 60(2):91-97
- 9: Shieh G. Exact Power and Sample Size Calculations for the Two One-Sided Tests of Equivalence. *PLoS One*. 2016 Sep 6;11(9):e0162093. doi: 10.1371 PMID: 27598468
- 
- 
-