Research Article

Open Access

Relative efficacy of virtual and in-office conservative care for musculoskeletal conditions

Holly Elliott, Erik Steven Moll, Eric Olmsted
Department of Bioinformatics and Population health analysis, Commonwealth Health Advisors, USA

Correspondence: Holly Elliott, Department of Bioinformatics and Population health analysis, Commonwealth Health Advisors, USA, Email holly.elliot@cwhadvisors.com

Introduction

Telemedicine has been increasingly popular recently, and it's potential utility in a wide variety of areas continues to increase as it becomes more popular with patients and more accessible to providers. In many areas, there is little doubt that the usability and functionality of telemedicine is equivalent to in-person care.[1] In physical therapy in particular, however, the potential of telehealth has been less than fully explored, as implementation has been slow due to payment and regulatory barriers.[2]

Patient satisfaction with telehealth, however, is high, and demand for this service is likely to increase.[3] 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. This is particularly of importance with respect to the ongoing aftereffects of COVID-19, which has left many in need of rehabilitative physiotherapy, with remote care being necessary to adequately address these needs.[4]

Patient satisfaction with telemedicine is high, and it is known to achieve a strong rate of visit completion.[5] However, data on its relative effectiveness is often lacking. Due to the short time frame during which virtual care options have been adopted, existing trials of rehabilitative physiotherapy have often provided inconclusive results or shown a high risk of bias.[6]

Given the urgency of need and the difficulty in carrying out long-term randomized clinical trials, it is therefore necessary to locate and apply statistical techniques with which to carry out this analysis despite the limited nature of the data.

We present a propensity score matching as a novel approach to resolving this issue as applied to a single program. This approach allows us to provide reasonably high-quality data, while minimizing the cost of conducting research as it does not require modification of the existing treatment of patients.[7] Propensity score matching achieves by controlling for known confounders, matching patients with patients who are similar in variables deemed likely to affect outcomes and thereby achieving results similar to a randomized controlled trial.[7]

Methods

The study was executed as a retrospective study on data collected from January 2019 to March 2021 as part of standard clinical operation. We compared a set of matched patient data to establish a comparative effectiveness of Airrosti Remote Recovery, Airrosti's virtual care model, as compared to in-office therapy. We limited our analysis to cases where all key variables were available. We controlled for confounding variables using propensity score matching within groups of the same area of injury (Figure 2 for groups). Using this matched data set, we evaluated the effectiveness of virtual conservative MSK care for upper and lower body injuries on pain improvement, surgical avoidance, injury resolution, and visit completion.

Patient waterfall

We identified patients from within Airrosti's internal database of all patients seen in January 2019 through March 2021. We then limited this group to those whose listed insurer was Blue Cross Blue Shield of Texas, in order to ensure access to insurance claims data for later analysis. Within those patients whose injury was a weight-bearing location (e.g, lower body), we then eliminated patients whose records did not contain useable BMI data. The patient counts are identified in Figure 1 below (Figure 1).
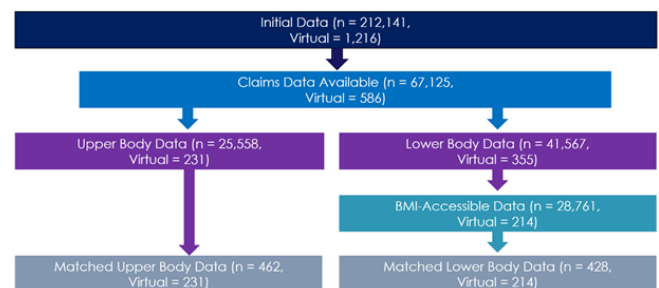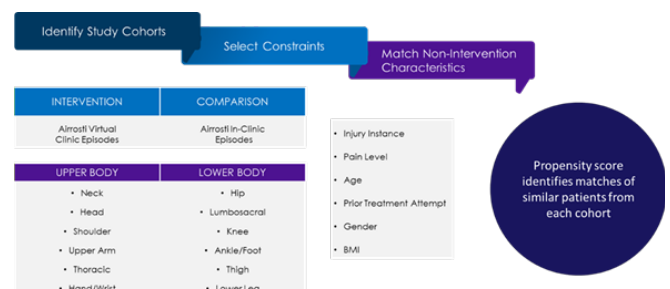

Figure 1 Patient Episodes.

Figure 2 Below provides an overview of our modeling approach.

Data

We obtained data from Airrosti's RainTree EMR system. These data contain patient diagnosis, treatment dates, comorbidities, limited data on social determinants of health, and the results of patient surveys before and after receiving Airrosti care.

Ethical considerations

Airrosti waived IRB approval for this study, as it was carried out as a retrospective study on previously existing, deidentified data. It is therefore outside the definition of human subjects research. The research presented in this article is designed for quality assurance and quality management, and in this context does not adversely affect rights or the welfare subjects. The discharge survey analysis would be performed regardless of the research, and no study results could affect clinical decisions about the patient's care, as the study was carried out well after care was concluded.

Categorizing injury location

Due to a preponderance of evidence that injuries in different locations have different outcomes,[8] we only permitted virtual subjects to match with in-office subjects 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).

Outcome variables

We evaluated patient outcomes using variables selected from prior reviews of the subject.[8]

These outcomes were:

I. Pain Improvement: The difference between initial and final reported pain on a 5-point scale,

II. Visit Completion: The absolute number of Airrosti visits completed,

III. Surgery Avoidance: Whether the patient reported avoiding a considered or scheduled surgery based on their Airrosti results,

IV. Injury Fixed: Whether the patient reported their injury as 'fixed' in post-therapy surveys (Figure 2).

Matching

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). We accomplished this matching using the MatchIt R package implementing the RELAX-IV algorithm via optmatch.[9] We selected these variables based on available data and characteristics controlled for in prior studies of musculoskeletal injury.[10] This matching approach yielded a high-quality matched dataset with an average standardized mean difference across all categories of 0.31, with all standardized mean differences being below 0.5.[11]

Single parametric tests (t-tests) are used to compare the significance of differences between two matched datasets.[11] In this study we want to examine the level of equivalence and answer the question "Are these two treatment modes similar or dissimilar?". We therefore applied a Two One-sided T-Test (TOST) approach to the matched data, which determines the maximum possible degree and direction of the difference in outcome between two conditions.[12]

Without having access to all conceivable data, it is impossible to prove that the effect of two conditions is 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 effect size and level of significance ($\alpha = .05$), there is no statistically meaningful difference between the two treatment modes (Figure 3).[13]

Variable	Upper Body Result	Lower Body Result
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

Figure 3 TOST Results.

TOST results

A Two One Sided T Test applies a one-sided parametric T-test to each side of the mean, establishing possible overall variance. Figure 3 summarizes the findings for each outcome variable. See figure 4 and 5 for precise 95% confidence intervals, shown as error bars. The upper and lower bounds of these figures represent a difference of plus or minus one half a standard deviation.
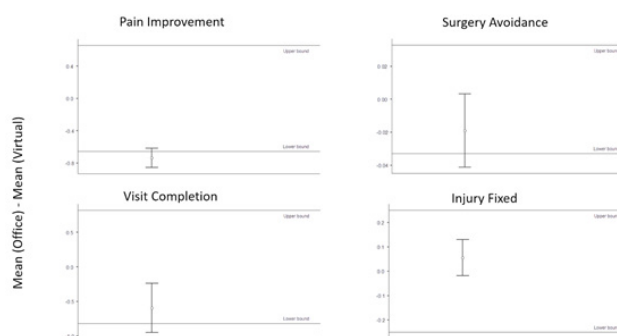

Figure 4 Upper Body TOST Results.

Negative values represent improved performance for the Virtual group. We observed improved perform

Negative values represent improved performance for the Virtual group. We observed improved performance for Virtual in visit completion and pain improvement (P = .02, P = .04). Surgical avoidance and injury resolution both fell below the upper bound of 1/2sd (P = .02, P = .03), indicating non-inferior performance ((Figure 5)).
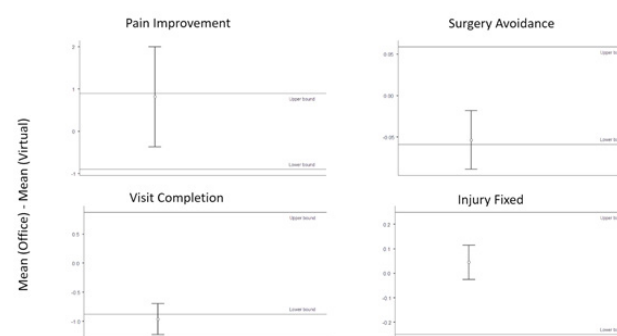

Figure 5 Lower Body TOST Results.

Negative values represent improved performance for the Virtual group. We observed improved performance for visit

Negative values represent improved performance for the Virtual group. We observed improved performance for visit completion and

surgical avoidance (P = .01, P = .04), and equivalent performance for injury resolution (P = .03 with respect to the upper bound). Results for pain improvement were ambiguous (Discussion).

Discussion

Principal Findings

The improvement in visit completion is an expected and validating result, as virtual care has been previously noted to improve physical therapy completion.[14] It is likely that this increased number of visits is the driving force behind the higher performance in pain improvement.

With respect to this patient group, the upper body results indicate promising success in remote treatment of these injuries. Although the higher performance in pain improvement is counterintuitive, it makes sense in light of the improvement in visit completion. This suggests that any downside imposed by the difficulties of remote treatment is made up for by the improved ability of patients to actually complete their course of physical therapy.

The lower body category (as shown in Figure 5) shows very distinct results compared to upper body with respect to pain improvement. The difference may fall within the equivalence range (P = .48). It is noteworthy that the confidence range is drastically wider than that of other variables.

Comparison with prior work

Prior work in this area typically involves high-cost randomized controlled trials.[1] This methodology offers comparatively greater ease of execution, while providing results that are sufficient to validate the efficacy of remote physical therapy for programs where no major controversy about the efficacy exists.

Limitations

The propensity score matching approach's principal limitation is the presence of untracked variables. We can match on a wide variety of elements, which are likely to at least function as correlated proxies for some untracked variables.[7] However, there will inevitably be differences between the patients who choose remote care vs. those who choose in-person care.

This methodology should not replace randomized controlled trials in remote physiotherapy evaluation, especially for areas where results are inconclusive. Given the wide standard deviation in lower-body care results, for example, it would be appropriate for programs seeking to evaluate lower-body remote physiotherapy to use randomized controlled trials.

Conclusion

These results suggest that the methodology used can effectively evaluate virtual care without a randomized clinical trial. This is supported by the validating findings on physical therapy completion, as well as matching quality measures. Together, these indicate that the typical conditions of remote physical therapy generate a good use case for propensity score matching, and that this analysis can reasonably be repeated.

The unusually broad standard deviation of pain reduction in the lower body group, combined with the similar performance as upper body in injury resolution and surgery avoidances, suggests that lower body injuries could be split into multiple categories, some of which will and will not be apt to respond to virtual conservative MSK care with respect to pain improvement. Further research could identify which subgroups are well-suited for particular treatment options, allowing for improved recommendations to patients seeking care for lower body musculoskeletal injuries.

Conflicts of interest

We declare there are no conflicts of interest.

Referenes

1. Colbert GB, Venegas–Vera AV, Lerma EV. Utility of telemedicine in the COVID–19 era. *Rev Cardiovasc Med*. 2020;21(4):583–587.

2. Lee AC, Davenport TE, Randall K. Telehealth Physical Therapy in Musculoskeletal Practice. *J Orthop Sports Phys Ther*. 2018;48(10):736–739.

3. Miller MJ, Pak SS, Keller DR, Barnes DE. Evaluation of Pragmatic Telehealth Physical Therapy Implementation During the COVID–19 Pandemic. *Phys Ther*. 2021;101(1):pzaa193.

4. Sun T, Guo L, Tian F, et al. Rehabilitation of patients with COVID–19. *Expert Rev Respir Med*. 2020;14(12):1249–1256.

5. Tenforde AS, Borgstrom H, Polich G, et al. Outpatient Physical, Occupational, and Speech Therapy Synchronous Telemedicine: A Survey Study of Patient Satisfaction with Virtual Visits During the COVID–19 Pandemic. Am J Phys Med Rehabil. 2020;99(11):977–981.

6. Seron P, Oliveros MJ, Gutierrez–Arias R, et al. Effectiveness of Telerehabilitation in Physical Therapy: A Rapid Overview. Phys Ther. 2021;101(6):pzab053.

7. Williamson EJ, Forbes A. Introduction to propensity scores. *Respirology*. 2014;19(5):625–635.

8. Schneider SP. Musculoskeletal injuries in construction: a review of the literature. *Appl Occup Environ Hyg*. 2001;16(11):1056–1064.

9. Ho DE, Imai K, King G, et al. "MatchIt: Nonparametric Preprocessing for Parametric Causal Inference." *Journal of Statistical Software*. 2011;42(8):1–28.

10. Peat G, McCarney R, Croft P. Knee pain and osteoarthritis in older adults: a review of community burden and current use of primary health care. A*nn Rheum Dis*. 2001 60(2):91–97.

11. Andrade C. Mean Difference, Standardized Mean Difference (SMD), and Their Use in Meta–Analysis: As Simple as It Gets. *J Clin Psychiatry*. 2020;81(5):20f13681.

12. Jankowski KRB, Flannelly KJ, Flannelly LT. The t–test: An Influential Inferential Tool in Chaplaincy and Other Healthcare Research. *J Health Care Chaplain*. 2018;24(1):30–39.

13. Shieh G. Exact Power and Sample Size Calculations for the Two One–Sided Tests of Equivalence. *PLoS One*. 2016;11(9):e0162093.

14. Herbert MS, Afari N, Liu L, et al. Telehealth Versus In–Person Acceptance and Commitment Therapy for Chronic Pain: A Randomized Noninferiority. *Trial J Pain*. 2017;18(2):200–211.