Scholten Treatment Effects Summary

Anthony Cunningham

Spring 2019

1 Introduction

Sandra Scholten is a student in the University of Iowa College of Nursing, pursuing a Doctorate of Nursing Practice. One requirement of this degree is the implementation and analysis of some intervention that addresses a problem that each student identifies. They then apply their intervention to a participating medical facility and analyze the effects of the intervention.

Sandra was concerned with the low rates of screening and treatment of osteoporosis at her participating facility in Des Moines. In an attempt to address this, Sandra implemented the Fracture Liaison Service (FLS) model of care by presenting this model to the orthopedic staff at her clinic. The osteoporosis screening and treatment rate outcomes were then examined before the FLS treatment was implemented, as well as after.

Sandra's objectives for her study were threefold:

1) Increase the referral rate of patients identified to have high risk of osteoporosis.
2) Increase the treatment rate of patients with osteoporosis-related fractures.
3) Increase staff knowledge relating to recommended osteoporosis referal and treatment protocols, as specified in the FLS model of care.

The purpose of this report is to present our findings regarding whether or not Sandra's intervention (the orthopedic education event) achieved her desired objectives, as well as provide visual representations of these results so that she may present these findings.

2 Data Collection

To evaluate her treatment in satisfying her first and second objectives, Sandra monitored the total number of patients who underwent initial evaluation for osteoporosis in two time periods: October through December of 2017, and October through December 2018. Sandra implemented her treatment prior to the 2018 time period, allowing a clear pre-implementation versus post-implementation distinction to be made. Within this sample of patients who underwent initial evaluation, Sandra collected data on how many patients were referred to an orthopedic specialist for further evaluation, as well as how many patients were treated for osteoporosis-related fractures. Sandra wished to observe significant increases in both the referral and treatment rates from the pre-implementation period to post-implementation.

To evaluate her treatment in satisfying her third objective, Sandra administered a test to the five staff members at her participating clinic. The test consisted of eight questions, and the same test was administered before implementation of her treatment as well as after implementation, creating natural pairing between each pre- and post-test. A staff member's knowledge is simply the number of correctly-answered questions on their test, with each score being an integer between 0 and 8.

2.1 Issues to Consider

One issue with Sandra's study is that it did not include a randomization procedure to detemine which observational unit, or which staff member, would receive the treatment, and which observational unit would not. As such, we cannot claim that any differences in referral rates or treatment rates were caused by her intervention, as there may have been some time effect in which (for example) referral rates increased due to changes in the orthopedic staff, in which these new staff members doled out referrals more liberally than their predecessor. We can only make a valid claim that, assuming a statistically significant difference exists, that a difference is present. We cannot claim that this difference was caused due to Sandra's intervention.

Another issue to consider is that Sandra did not collect data on which staff member made the determination on whether or not to refer or treat a patient. Since there are only five orthopedic staff members, it may be that each staff member had differing philosophies on referring and treating patients. If so, each staff member's impact on the overall rates would be great, especially with the small number of staff members. This clustering is not taken into account, potentially biasing our results, although it is difficult to say in which direction this bias would occur, as some staff members may be more cautious is treating or referring patients, while some take more chances during diagnosis.

A third issue is the small number of patients seen in each time period, as well as the small number of staff members who completed Sandra's test. Generally, small samples affect the power of analyses of differences; only big differences are judged to be statistically significant. For the former, this manifests into small cell counts in their respective 2-by-2 contingency tables. In order to mitigate this small sample effect, our analysis of comparing the differences in proportions was done using Fisher's Exact Test. We analyzed the test data using a non-parametric test: the Wilcoxon-Pratt Signed-Rank Test. This is analogous to conducting a paired t-test. However, with only five pairs of observations, we determined that a t-test would not be valid to conduct.

3 Methodology

In examining the success of Sandra's first and second objectives, we conducted a Fisher's Exact Test to detect significant differences in referral and treatment rates pre- versus post-implementation. A Fisher's Exact Test calculates a p-value directly from a probability distribution function, the hypergeometric distribution. This is due to our small-sample limitation, as using large-sample methods for obtaining statistical significance are unreliable in this study. In the original form of the test, our data is constructed into a 2-by-2 contingency table, with each cell corresponding to each unique combination of treatment or referral distinction and time period, while the numbers in these cells correspond to the frequency with which we observed each combination. The key assumption is that the marginal frequencies are fixed; they do not change (Gibbons [1]). One result of this is that knowing the count in one cell allows one to know the counts in all other cells. This allows for a simpler computation of the p-value: we calculate the probability that we would observe a count in one cell as extreme or more extreme under the null hypothesis that the two factors are independent: knowing the value of one factor will not inform the value of the other factor. To calculate this probability, we use the hypergeometric distribution (Gibbons [1]).

In order to examine the success of Sandra's third objective, we conducted a non-parametric Wilcoxon-Pratt Signed-Rank Test to detect significant differences in staff test scores pre- versus post-implementation. This test is analogous to a paired t-test in that we are testing whether or not each difference within pairs is significantly different from zero; that there was an observed change. However, utilization of the t-test would be unreliable due to our small sample size of 5 staff members, which contradicts a key assumption that our differences within pairs are approximately normally-distributed. We use non-parametric test because these have less stringent distributional assumptions. For the Wilcoxon-Pratt Signed-Rank Test, it is only assumed that each difference within pairs is continuous and symmetric about the median (Randles [3]). The Signed-Rank Test calculates the probability that we observed these differences under the null hypothesis that no change in osteoporosis treatment and referral knowledge is present. To calculate, we first calculate the test statistic, W. First, we calculate each difference within pairs. Then, we rank the absolute differences from smallest to largest. The Pratt method is used to handle ties, or differences of zero. In the original Wilcoxon test, ties are thrown out before ranking. Pratt's method includes these ties while ranking, and then throws them out prior to summing up the ranks (Pratt [2]). For negative differences, we multiply their respective ranks by -1, and then we add up all of the positive and negative ranks. This constitutes our test statistic, W. With our small sample, we calculate an exact p-value, where R calculates the asymptotic approximation of the null distribution and calculates the p-value from that.

4 Results and Graphical Output

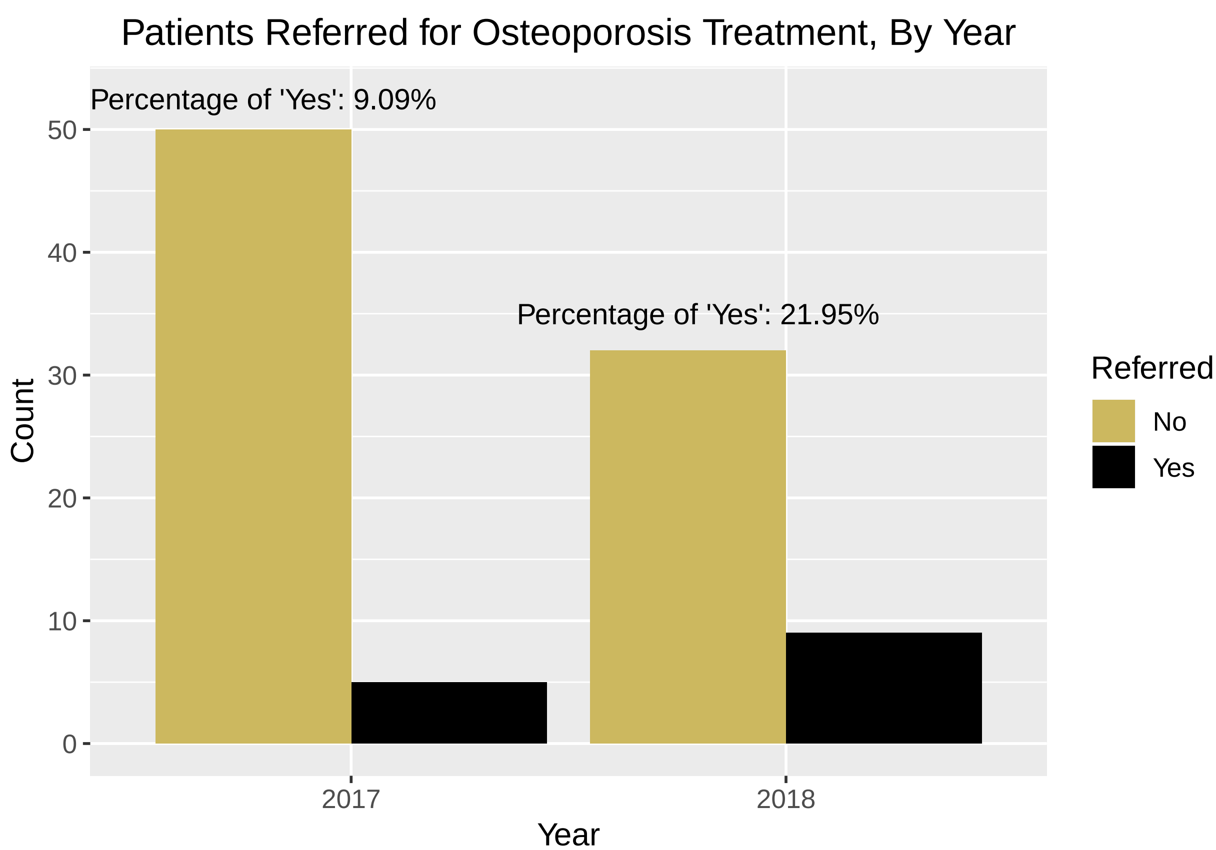
4.1 Objective 1: Increase the Referral Rate for High-Risk Patients

4.1.1 Referral Counts and Rates Before vs After

```
##              2017 2018 Total
## Referred        5    9    14
## Not Referred   50   32    82
## Total          55   41    96
```

```
##                          2017 2018
## Proportion Referred     0.091 0.22
## Proportion Not Referred 0.909 0.78
```

4.1.2 Graphical Output



4.1.3 Significance Test: Fisher's Exact Test

```
##
##  Fisher's Exact Test for Count Data
##
## data:  referred
## p-value = 0.08861
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##   0.7560782 11.5672308
## sample estimates:
```

```
## odds ratio
##   2.781442
```

4.1.4 Conclusion

According to Fisher's Exact Test, we have some evidence of a statistically significant difference in the proportion of patients evaluated for osteoporosis who were referred to an orthopedic specialist after intervention vs. before intervention (p-value = 0.08861). We are 95% confident that the odds of an evaluated patient being referred was between 0.756 and 11.567 times greater after intervention than before. This result is very encouraging because, even with the small sample considered, we were still able to detect a significant increase in the referral rate (at a 10% error rate). Furthermore, a greater proportion of patients obtained consent for further evaluation of a devastating disease, allowing them the opportunity to obtain the information necessary for orthopedic specialists to diagnose and treat this disease. While we are not able to say that Sandra's intervention caused this, we can point to this result as a reason to study this FLS model of care further in a randomized experiment.

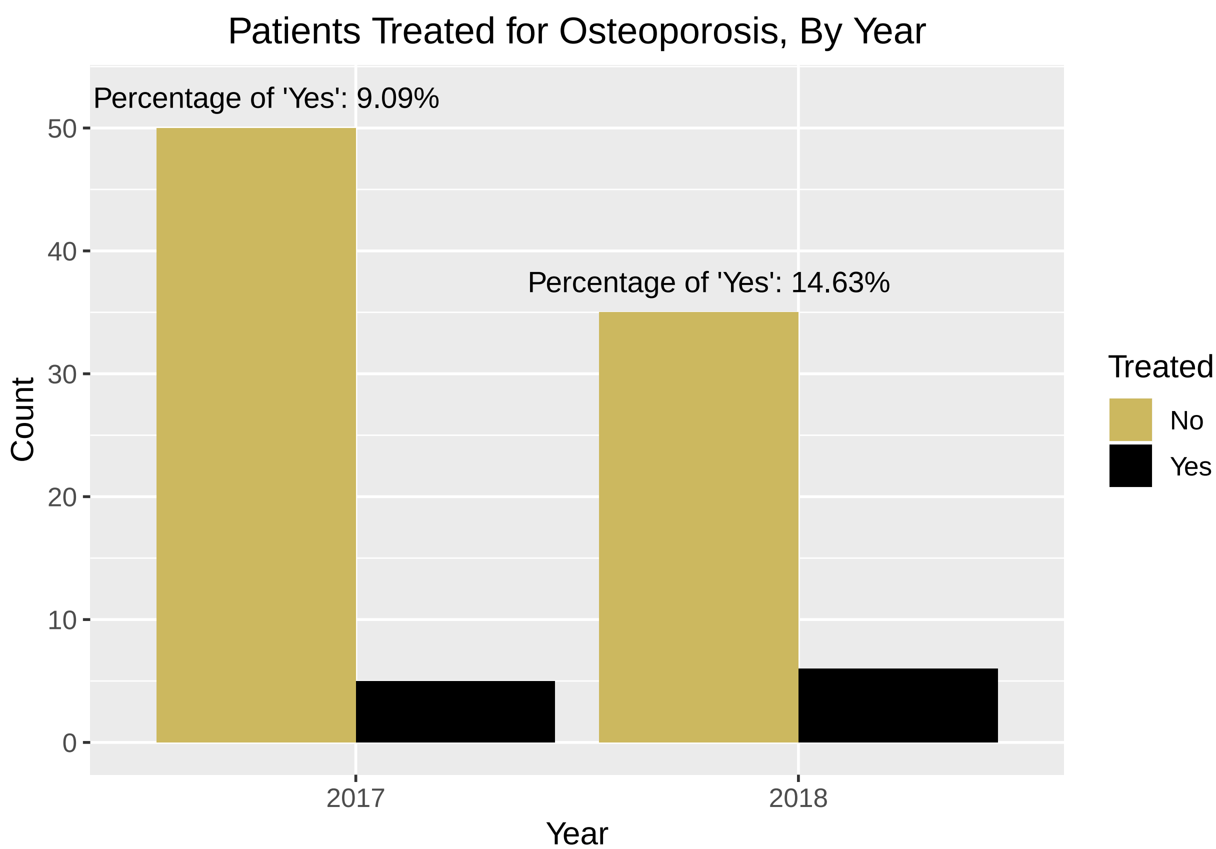
4.2 Objective 2: Increase the Treatment Rate of Patients with Osteoporosis-Related Fractures

4.2.1 Treatment Counts and Rates Before vs After

```
##               2017 2018 Total
## Treated          5    6    11
## Not Treated     50   35    85
## Total           55   41    96
```

```
##                         2017  2018
## Proportion Treated     0.091 0.146
## Proportion Not Treated 0.909 0.854
```

4.2.2 Graphical Output



4.2.3 Significance Test: Fisher's Exact

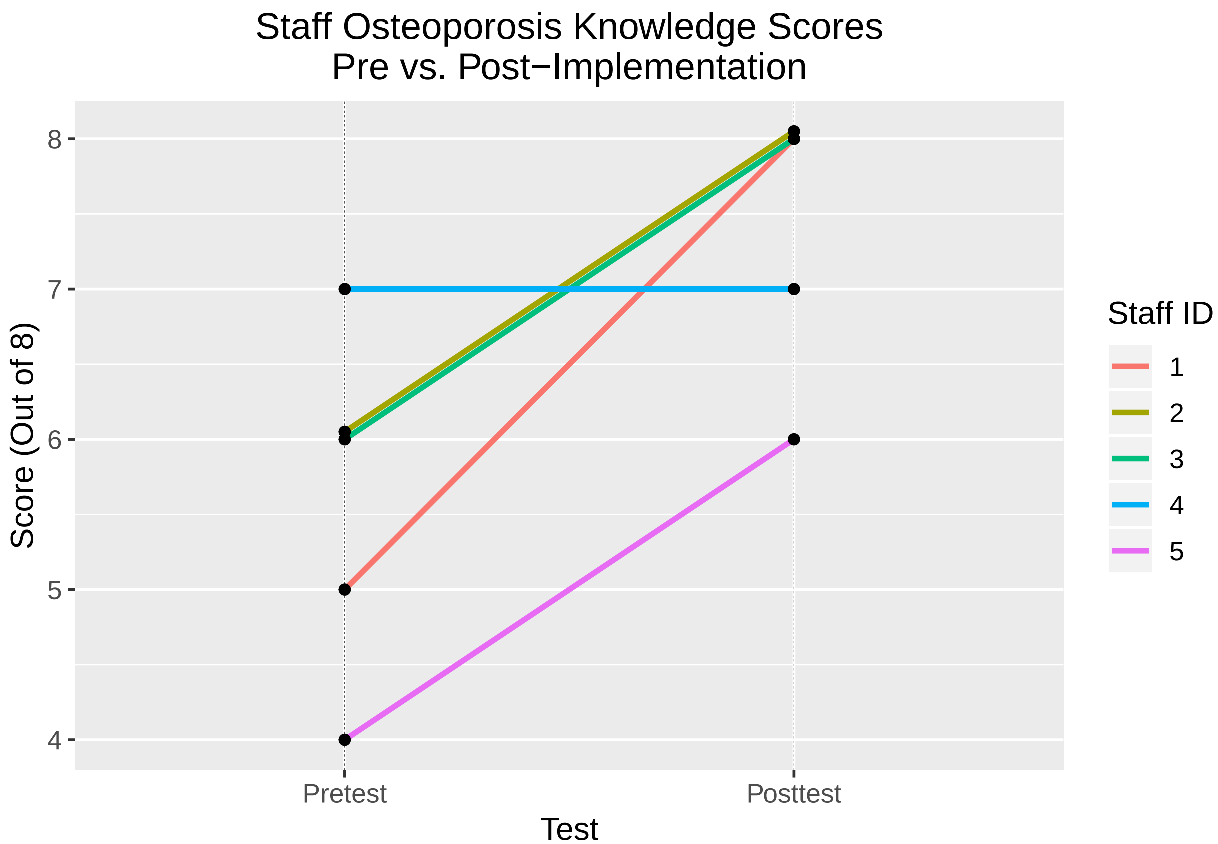
```
##
##  Fisher's Exact Test for Count Data
##
## data:  treated
## p-value = 0.5206
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##  0.3983376 7.6607763
## sample estimates:
## odds ratio
##   1.704437
```

4.2.4 Conclusion

According to Fisher's Exact Test, we do not have evidence of a significant difference in the proportion of patients getting treated for osteoporosis-related fractures after intervention vs before (p-value = 0.5206). While this result is disappointing, we can use this opportunity to address the issues in study design and sample size. Due to a lack of a randomization of the treatment on our subjects, we cannot establish a causal inference between the treatment and the results. Additionally, the small samples do not give us much statistical power to determine significant differences before and after treatment. In this case, the difference was not large enough to supply us with evidence of a significant difference.

4.3 Objective 3: Increase Staff Knowledge of Recommended Osteoporosis Referral and Treatment Protocols

4.3.1 Graphical Output



4.3.2 Significance Test: Wilcoxon-Pratt Signed-Rank Test

```
##
##  Exact Wilcoxon-Pratt Signed-Rank Test
##
## data:  y by x (pos, neg)
##   stratified by block
## Z = 1.9415, p-value = 0.125
## alternative hypothesis: true mu is not equal to 0
```

4.3.3 Conclusion

According to the Wilcoxon-Pratt Signed-Rank Test, we do not have enough evidence to show a statistically significant difference in knowledge of recommended osteoporosis referral and treatment protocols, after accounting for staff member (p-value = 0.125). This result is frustrating, as no staff member scored worse post-implementation, while four out of five staff members scored better than before. Empirically, we see obvious improvements in knowledge among staff members, but with only five pairs of observations, we have little statistical power with which to claim a statistically significant difference. However, I would feel confident that if Sandra's treatment were to be applied on a larger scale, we would observe a significant increase in staff knowledge.

5 Final Remarks

Despite the design and sample size issues of this study, Sandra should be excited about these results. Even if we cannot establish a causal effect of Sandra's treatment and the increase in proportion of patients receiving referrals and the proportion of patients getting treated by an orthopedic specialist, we did see improvements in these rates (even if they are not statistically significant, in the case of the latter analysis), meaning that a greater proportion of patients were seen and treated by a specialist, affording more high-risk patients the opportunity to detect osteoporosis, as well as mitigate these potential risks in a proactive fashion. Our results at least keep the door open for further exploration of Sandra's treatment, in a randomized environment on a larger scale. This could stand to benefit a greater number of patients if further researched.

6 References

Dickinson Gibbons, J. (2006). Fisher's Exact Test. In Encyclopedia of Statistical Sciences (eds S. Kotz, C. B. Read, N. Balakrishnan, B. Vidakovic and N. L. Johnson). doi:10.1002/0471667196.ess0791.pub2

Pratt, John (1959). Remarks on Zeros and Ties in the Wilcoxon Signed Rank Procedures. In Journal of the American Statistical Association, 54:287, 655-667, doi: 10.1080/01621459.1959.10501526

Randles, R. H. (2006). Wilcoxon Signed Rank Test. In Encyclopedia of Statistical Sciences (eds S. Kotz, C. B. Read, N. Balakrishnan, B. Vidakovic and N. L. Johnson). doi:10.1002/0471667196.ess2935.pub2