Journal Club Questions

JPAG Article, June 2019: Dowlut-McElroy T, et al. *Treatment of Prepubertal Labial Adhesions: A Randomized Controlled Trial.*

1. This study is a randomized control trial. Do you think this was a well-designed study? What are the study design types of randomized controlled trials?

Q1 Answer Key

- Strengths:
    - Randomization and medication was blinded to PI, study staff and participants (double blinded)
    - Block randomization used so that for each 10 subsequent participants enrolled, there was an equal number of participants randomized to each of the 2 treatment groups
    - Examiners were trained on the use of severity and thickness
    - Thorough inclusion/exclusion criteria
    - Authors met their minimum number of patients per group based on power analysis

- Weaknesses:
    - authors use of a clinically meaningful rate failure rate for the power analysis (50% vs 10% at 3 weeks). Other investigators may have chosen a different clinically meaningful failure rate as this is a subjective measure.
    - Did not control for adherence
    - No objective measure of traction

- Types of RCT
    - Parallel Group
    - Factorial
    - Cross over
    - Cluster

2. The authors report that the relatively small sample size likely contributed to the lack of statistical significance between the 2 treatment groups (estrogen vs emollient). What 4 components are necessary to calculate sample size for trials with dichotomous outcomes (i.e. complete resolution of labial adhesion or not)? Did the authors provide all the necessary assumptions for readers to replicate their sample size calculation? Go to https://clincalc.com/stats/samplesize.aspx (or another sample size calculator). Based on the info they provided did they have enough patients in each group?

Q2 Answer Key

- type I error (α), power, event rate in the control group, and a treatment effect of interest (or an event rate in the treatment group). Can also calculate an effect size from the two event rates
- Yes (0.05, 80%, 50%, 10%)
- Sample size 19 participants per group, total of 38 participants

- Yes they had enough in each group based on calculation

3. Although the results were not statistically significant, the estrogen intervention group was 2X as successful as the Cetaphil group (36% vs. 19%) with respect to complete resolution. Though not statistically significant likely due to low sample size, would you consider the results to be clinically significant? How, if at all, would you incorporate these findings into your clinical practice? How would you use their results in a future study?

-

Q3 Answer Key

- The authors of this study chose what they deemed to be a clinically meaningful change: 10% resolution vs 50% resolution. This large difference (40%), which is clinically significant, resulted in the calculation of a small sample size and likely resulted in lack of statistical significance. What they found in the study is that the difference in complete resolution labial adhesions was actually less than the 40% that they expected (36% vs 19% à 17%).
- The findings suggest that with time (with traction) there was an improvement in labial adhesions within both treatment groups. However, the magnitude of improvement was greater with the estrogen group. For families that choose not to use estrogen for fear of side effects, this study shows that labial adhesions can improve with the use of an emollient and traction albeit not as fast and not as much as with estrogen. Could it be that traction is key rather than estrogen? Yes, the lack of objective measurement of adherence is an issue which should be addressed in future studies to evaluate whether estrogen is needed for the treatment of labial adhesions.
- If this data is used (36% vs 19% in a post hoc analysis) to calculate a sample size, a total of 214 participants are needed to achieve statistical significance.
- A note on sample size:
    - A study with a small sample size and lack of statistical significance in the result is not poor or weak. What this study shows is the difference between clinical and statistical significance and that the use of clinical data may result in results that are not statistically significant. Although not statistically significant, clinically, the difference of 17% may not be that important to families who may choose just an emollient rather than estrogen given the potential side effects of estrogen.

4. The authors created a composite score to determine adhesion severity. What do you think about the variables included in this score? How do you interpret the significance of this score?

Q4 Answer Key

- The use of the composite score provides a more objective method of describing labial adhesions and makes comparison of groups easier (as compared to other manuscripts about labial adhesions).

- For labial adhesions, there are no familiar scoring systems on which to record outcomes. There has been no standard way of reporting labial adhesions in the past – previous authors have described adhesions as percentage closure over vaginal opening, length of involvement of labia, some 'filmy" or thin adhesions etc The authors utilized the composite score so they could

determine the effect size of treatment (i.e. variation in scores representative of labial adhesion reported as mean and SD).

- Variables of labial adhesions could include age at diagnosis of adhesion, duration of adhesion presence, history of perineal pain, history of UTI, prior treatments trialed on the adhesion and their success/failure, thickness of adhesion and length of adhesion

5. The authors identify a significant limitation in their study, ie: adherence to treatment was collected only by patient report and treatment was not observed. The authors state that the overall lower complete resolution rates of labial adhesions in this study might be a function of lack of adherence. Do you agree? Would there any way to mitigate this limitation as consistent observed therapy would be challenging to achieve?

Q5 Answer Key

- Yes, lack of adherence may have contributed to the lower rate of complete resolution in the estrogen group.
- Investigators could have weighed the medication container and asked the families to bring in the medication container at each visit. Weight pre- and post-visits would have been a more objective method to assess adherence.