Peer Review File

Article information: https://dx.doi.org/10.21037/tau-23-454

Reviewer A

Comment 1: It is a pity that patients receiving only clomiphene are not presented. However, there are some publications showing the results after such treatment and they should be included in the Discussion i.e. Huijben et al. (Andrology 2023).
Reply 1: We thank the reviewer for this comment and agree that including a more in-depth discussion of some of the clomiphene monotherapy data would be valuable context to add to our discussion.
Changes in the text:

Lines 179-185: "*Although our study did not include a clomiphene monotherapy arm, previous investigations have explored the impact of clomiphene on fertility-related outcomes. In a systematic review and meta-analysis (7) of studies investigating this question, Huijben et al. found that clomiphene significantly increases sperm concentration, motility, morphology, total testosterone, estradiol, and gonadotropin levels. There were no major adverse events reported. This meta-analysis should be interpreted within the context of the quality of data reported in the included studies; sample sizes, for example, varied from as low as 11 to as high as 140 participants.*"

Comment 2: The number of patients in both groups is low, so the results should be considered as preliminary.
Reply 2: Unfortunately, much of the literature on male factor infertility and associated treatments share this limitation. We have, in multiple parts of the manuscript, commented on the limited sample size; however, we agree with the reviewer and have added additional text to ensure there is no ambiguity on this.
Changes in the text:

Lines 208-209: "*Given the small sample size, results should be considered preliminary.*"

Comment 3: Dosage of clomiphene and anastrazole should be shown, as well as the length of therapy.
Reply 3: Clomiphene dosage was 25 mg QD or 50 mg QOD; this has been added to the manuscript. The dosage of anastrozole was not as consistent; this is already described in the results section of the manuscript and these data are outlined in Table 1. The length of therapy is included in the text of the results section as the days between pre- and post-treatment variables; given the retrospective nature of the study we are unable to provide a more accurate statement on treatment duration (cannot be guaranteed that participants started therapy immediately after it was prescribed or a week later, for example).
Changes in the text:

Lines 106-108: "*Anastrozole dosing was not standardized, and thus the median dose and interquartile range (IQR) were reported when baseline patient characteristics were considered. Clomiphene dosage was standardized to 25 mg QD or 50 mg QOD.*"

Lines 136-137: "*The average number of days between pre- and post-treatment variables was 144 for the anastrozole group compared to 109 for the combination group (p = 0.0548).*"

Comment 4: Adverse effects should be shown in the Results.
Reply 4: There were no major adverse effects reported. This has now been added to the results section.
Changes in the text:

Line 140: "*There were no major adverse events reported.*"

Comment 5: Normospermia should be replaced by normozoospermia.
Reply 5: We have made this change. Thank you.
Changes in the text: All references to normospermia have been changed to normozoospermia.

Reviewer B

Comment 1: Abstract: clarify infertility definition and explain the aim of the research (Improve semen parameters?). Explain why you said "modest benefits"
Reply 1: We have clarified that this study includes participants with both primary and secondary infertility. We have provided additional language to clarify the aim of the study. The word modest has been removed; we agree with the reviewer that this language at best does not add detail to the abstract and at worst is confusing to the reader.

Changes in the text: The abstract has been edited, as noted above.

Line 36-51: "*In men with impaired semen parameters, empiric medical therapies such as clomiphene citrate, a selective estrogen receptor modulator, and anastrozole, a selective aromatase inhibitor, are often employed. The effects of jointly administering these agents on semen parameters are not well understood. Here, we describe the findings of our multi-center, retrospective cohort study of men with idiopathic primary or secondary infertility. N=21 men were treated with combination therapy (anastrozole and clomiphene) and N=69 men were treated with monotherapy (anastrozole). Patients with pre-treatment normozoospermia and recent or current exogenous testosterone therapy were excluded. Baseline and post-treatment semen and sex hormone parameters were compared among groups. Following treatment, 43% of men in the combination therapy group demonstrated normozoospermia, compared to 25% in the monotherapy group. Furthermore, men in the combined group demonstrated marked improvements in total motile sperm count (11.3 vs. 2.1 M, p = 0.03). There were no significant differences in hormone levels among the two groups following treatment. Combination therapy with clomiphene citrate and anastrozole was associated with modest benefits in post-treatment semen parameters, when compared to anastrozole monotherapy. Future investigations with larger sample sizes and prospective study designs are necessary.*"

Comment 2: Introduction: clarify the objective or aim of the project
Reply 2: We have added clarifying language.
Changes in the text:

Lines 70-72: "*Here, we leverage the same multi-institutional cohort of patients to query whether there is additional benefit in combination therapy with clomiphene and anastrozole, compared to anastrozole alone with respect to semen parameters.*"

Comment 3a: Semen analysis protocol, how many samples were collected, which semen parameters were collected, methodology used to evaluate semen analysis
Reply 3a: Additional text has now been added to address this comment.
Changes in the text:

Lines 91-97: "*Quantitative semen analysis was performed according to World Health Organization 5th Edition criteria using a computer assisted semen analyzer (Medical Electronic Systems, Encino, CA). Samples demonstrating oligozoospermia or azoospermia were independently evaluated using high-powered microscopy. Semen volume, concentration, total motility, strict morphology (Kruger), and pH were obtained. All patients included in the study underwent two semen analyses prior to starting therapy, and the most recent sample was used for pre-intervention data.*"

Comment 3b: Clarify the type of laboratory/methodology used for the labs included in this study.
Reply 3b: Additional text has now been added to address this comment.
Changes in the text:

Lines 112-113: "*Both institutions used electrochemiluminescence immunoassays to quantify hormone levels using standard clinical laboratory practices.*"

Comment 3c: If 2 different labs were used to evaluate estradiol levels. Are these comparable?
Reply 3c: All participant from CCF had labs drawn and processed by CCF; all participants from UCLA had labs drawn and processed by UCLA. As noted in our response to Comment 3b, both labs adhere to the highest quality of standard clinical laboratory practices. The two estradiol labs differed with respect to their cutoff and this is addressed in the text already present in the methods section and also, more broadly, in the limitations section of our text.
Changes in the text:

Lines 113-116: "*Estradiol assays used by each institution differed with respect to the lowest level of detectable hormone (CCF: 25 pg/mL, UCLA: 12 pg/mL). Therefore, baseline E2 concentration was treated as a binary categorical variable with a cutoff set at ≥25 pg/mL.*"

Lines 209-211: "*There are many benefits to a multi-institutional approach, such as the one we employ here; however, with such a collaboration comes potential inter-facility measurement bias for lab testing and SAs.*"

Comment 3d: How treatment was tracked, how adherence to the treatment was followed, which dose was prescribed for Anastrozole and for clomiphene
Reply 3d: Our report describes a retrospective study with the associated limitations of a retrospective study. The medications were well tolerated and no participants reported stopping taking the medications. This information has been added to the manuscript. We are unable to comment further regarding adherence. This is already encompassed to an extent in the limitations section of our manuscript, which outlines the limitations of such a retrospective, rather than prospective, study design. We have added additional text for further clarification. Clomiphene dosage was 25 mg QD or 50 mg QOD; this has been added to the manuscript. The dosage of anastrozole was not as consistent; this is already described in the results section of the manuscript and these data are outlined in Table 1.
Changes in the text:

Lines 106-108: "*Anastrozole dosing was not standardized, and thus the median dose and interquartile range (IQR) were reported when baseline patient characteristics were considered. Clomiphene dosage was standardized to 25 mg QD or 50 mg QOD.*"

Line 140: "*There were no major adverse events reported.*"

Lines 203-205: "*The medications were well-tolerated and no participants reported stopping the medications; however, treatment adherence was not queried beyond self-report.*"

Comment 3e: Any criteria to start clomiphene citrate
Reply 3e: Language addressing this comment has been included in the manuscript.
Changes in the text:

Line 84-88: "*Initiation of clomiphene therapy for off-label, empiric use in the context of hypogonadal symptoms and/or subfertility was determined after careful discussion and shared decision-making between physician and patient. In general, patients with symptomatic hypogonadism (low testosterone or low-normal testosterone) and/or impaired sperm parameters was offered treatment.*"

Comment 3d: If you described testicular volume in your results, could you explain how this was measured?
Reply 3d: Language addressing this comment has been included in the manuscript.
Changes in the text:

Line 88-89: "*Testicular volume was assessed by clinical examination and orchidometer if there were difficulties with clinical examination alone.*"

Comment 4a: Describe briefly demographic parameters, such as median age, comorbidities, median dose anastrozole, and clomiphene citrate, and median time FU post treatment.
Reply 4a: Language addressing this comment has been included in the manuscript. For additional detail, readers are invited to reference Naelitz et al., which leveraged the same dataset and includes further demographic data. This reference has been explicitly included in the results section and

several times in the manuscript as well. Furthermore, Table 1 includes this data stratified by treatment arm.
Changes in the text:

Lines 126-129: "*Ninety participants were included in this analysis; baseline characteristics of all participants are reported in our recently published study (4). In brief, the median age of participants was 36 (IQR 32-41), BMI 32 (27-43), and anastrozole dose 3 mg/wk (3-7). The median follow-up duration was 91 days (IQR 64-117).*"

Table 1.

Comment 4b: Median time between initial and second semen analysis
Reply 4b: Language addressing this comment has been included in the manuscript.
Changes in the text:

Line 128-129: "*The median follow-up duration was 91 days (IQR 64-117).*"

Comment 4c: % of men with normal, oligo, severe oligo, and azoo, and how many or % of men with azoo, severe oligo and oligo improved the sperm count after treatment
Reply 4c: Language addressing this comment has been included in the manuscript
Changes in the text:

Line 129-131: "*With respect to baseline WHO semen classification categories, 19 participants (21%) demonstrated azoospermia, 11 (12%) cryptozoospermia, 32 (36%) severe oligozoospermia, and 28 (31%) oligozoospermia.*"

Line 149-154: "*With respect to any improvements in TMSC, the anastrozole monotherapy group, demonstrated the following results stratified by pre-treatment WHO category: azoospermia, 2 participants (13%) with improved TMSC; cryptozoospermia, 4 participants (44%); severe oligozoospermia, 16 (64%); and oligozoospermia, 11 (55%). For the combined group: azoospermia, 0 participants (0%) with improved TMSC; cryptozoospermia, 1 participant (50%); severe oligozoospermia, 7 (100%); and oligozoospermia, 6 (75%). This was not statistically significant.*"

Comment 4d: Did motility improve with the treatment?
Reply 4d: This information is outlined in Table 1. The text in the results section does explicitly comment about the improvements in TMSC. In an effort to enhance comprehensibility and not bloat the results section, we feel that additional text that repeats what is outlined in the table is not optimal.
Changes in the text: No changes in the text were necessary given the content of Table 1 and the results section.

Comment 4e: % of men with low T.
Reply 4e: This information is outlined in Table 1. In an effort to enhance comprehensibility and not bloat the results section, we feel that additional text that repeats what is outlined in the table is not optimal.

Changes in the text: No changes in the text were necessary given the content of Table 1 and the results section.

Comment 4f: % men who improved T levels after treatment, % men with low T who improved T levels after treatment
Reply 4f: The focus of this report is on improvements of sperm parameters following either monotherapy with anastrozole or combination therapy with anastrozole and clomiphene. We have added additional text to the introduction and abstract to make this more explicit. Previous reports have discussed the impact of these agents on T levels. Given this, the authors do not feel it adds to the manuscript to outline these results in detail beyond the sufficient detail that is presented in Table 1 and what can be referenced in Naelitz et al., which leveraged the same dataset.
Changes in the text: No changes in the text were necessary given the above response.

Comment 5: Several limitations should be included, especially labs for semen analysis and hormones, data collection, and follow-up.
Reply 5: Language addressing this comment has been included in the manuscript.
Changes in the text:

Lines 197-215: "*Our study is not without limitations. In this report, we present real-world, multi-institutional data, which was retrospectively collected. For this reason, the two comparison groups have innate differences that are not controlled for, limiting the generalizability of our findings. Other factors that may limit generalizability include our participants' demography and baseline characteristics. Of note, the median BMI of both groups is greater than 30, which may impact the efficacy of medications such as anastrozole and results may be different if higher doses were used or if this study were repeated on a patient population with a lower BMI. The medications were well-tolerated and no participants reported stopping the medications; however, treatment adherence was not queried beyond self-report. Most relevant are the pre-treatment levels of LH and the associated T:LH ratio, which differ among the two groups. Furthermore, given the limited sample size of men on combination therapy, we are unable to perform higher level statistical analyses such as multivariate regression with a high degree of fidelity. Given the small sample size, results should be considered preliminary. There are many benefits to a multi-institutional approach, such as the one we employ here; however, with such a collaboration comes potential inter-facility measurement bias for lab testing and SAs. Even with a multi-institutional approach, our sample size remains relatively small—underscoring the innate challenges of studying this patient population. Future investigations with prospective study designs and a more robust number of participants may benefit from higher level statistical methodology that is not possible in this report, including propensity match score analysis.*"

Reviewer C

Comment 1 and 2: The dose dependency analysis is necessary for developing a new treatment. We will consider that post-treatment T level will be significantly increased if the dose of anastrozole is increased, if the treatment duration is extended. Difference of T/LH between the two groups is disastrous. Propensity
much score analysis will be necessary in this case but it needs larger number of participants.
Reply 1 and 2: These are excellent points. Our study is limited by its retrospective design and sample size so additional statistical analyses would not be fruitful; however, it represents an important jumping off point for future investigations. With respect to the comment about differences in T/LH, we recognize that this is a major weakness of this study and highlight it throughout the manuscript – an unfortunate consequence of analyzing real world data.
Changes in the text: We have expanded on our limitations section to highlight the limitations inherent with a retrospective study on a relatively limited number of participants.

Lines 197-215: "*Our study is not without limitations. In this report, we present real-world, multi-institutional data, which was retrospectively collected. For this reason, the two comparison groups have innate differences that are not controlled for, limiting the generalizability of our findings. Other factors that may limit generalizability include our participants' demography and baseline characteristics. Of note, the median BMI of both groups is greater than 30, which may impact the efficacy of medications such as anastrozole and results may be different if higher doses were used or if this study were repeated on a patient population with a lower BMI. The medications were well-tolerated and no participants reported stopping the medications; however, treatment adherence was not queried beyond self-report. Most relevant are the pre-treatment levels of LH and the associated T:LH ratio, which differ among the two groups. Furthermore, given the limited sample size of men on combination therapy, we are unable to perform higher level statistical analyses such as multivariate regression with a high degree of fidelity. Given the small sample size, results should be considered preliminary. There are many benefits to a multi-institutional approach, such as the one we employ here; however, with such a collaboration comes potential inter-facility measurement bias for lab testing and SAs. Even with a multi-institutional approach, our sample size remains relatively small—underscoring the innate challenges of studying this patient population. Future investigations with prospective study designs and a more robust number of participants may benefit from higher level statistical methodology that is not possible in this report, including propensity match score analysis.*"

Comment 3: BMI of both groups is more than 30, which is not normal in world wide. This also tell us the possibility that the dose of anastrozole is insufficient.
Reply 3: Language addressing this comment has been included in the manuscript, mainly the limitations section.
Changes in the text: We have expanded on our limitations section.

Lines 197-215: "*Our study is not without limitations. In this report, we present real-world, multi-institutional data, which was retrospectively collected. For this reason, the two comparison groups have innate differences that are not controlled for, limiting the generalizability of our findings. Other factors that may limit generalizability include our participants' demography and baseline characteristics. Of note, the median BMI of both groups is greater than 30, which may impact the*

efficacy of medications such as anastrozole and results may be different if higher doses were used or if this study were repeated on a patient population with a lower BMI. The medications were well-tolerated and no participants reported stopping the medications; however, treatment adherence was not queried beyond self-report. Most relevant are the pre-treatment levels of LH and the associated T:LH ratio, which differ among the two groups. Furthermore, given the limited sample size of men on combination therapy, we are unable to perform higher level statistical analyses such as multivariate regression with a high degree of fidelity. Given the small sample size, results should be considered preliminary. There are many benefits to a multi-institutional approach, such as the one we employ here; however, with such a collaboration comes potential inter-facility measurement bias for lab testing and SAs. Even with a multi-institutional approach, our sample size remains relatively small—underscoring the innate challenges of studying this patient population. Future investigations with prospective study designs and a more robust number of participants may benefit from higher level statistical methodology that is not possible in this report, including propensity match score analysis."

Comment 4: Please indicate the patient background, study protocol, results, including adverse effects, in detail.
Reply 4: Additional information has been added throughout the methods, results and limitations section to address this comment. This is not a double-blind, placebo-controlled randomized control trial but rather a retrospective, cross-sectional study so further details are not possible beyond what is presented here.
Changes in the text: Language addressing this comment has been included in the manuscript, mainly the methods, results, and limitations sections.

Line 84-88: "*Initiation of clomiphene therapy for off-label, empiric use in the context of hypogonadal symptoms and/or subfertility was determined after careful discussion and shared decision-making between physician and patient. In general, patients with symptomatic hypogonadism (low testosterone or low-normal testosterone) and/or impaired sperm parameters was offered treatment.*"

Line 88-89: "*Testicular volume was assessed by clinical examination and orchidometer if there were difficulties with clinical examination alone.*"

Lines 106-108: "*Anastrozole dosing was not standardized, and thus the median dose and interquartile range (IQR) were reported when baseline patient characteristics were considered. Clomiphene dosage was standardized to 25 mg QD or 50 mg QOD.*"

Line 140: "*There were no major adverse events reported.*"

Lines 203-205: "*The medications were well-tolerated and no participants reported stopping the medications; however, treatment adherence was not queried beyond self-report.*"

Lines 197-215: "*Our study is not without limitations. In this report, we present real-world, multi-institutional data, which was retrospectively collected. For this reason, the two comparison groups have innate differences that are not controlled for, limiting the generalizability of our findings. Other factors that may limit generalizability include our participants' demography and baseline*

characteristics. Of note, the median BMI of both groups is greater than 30, which may impact the efficacy of medications such as anastrozole and results may be different if higher doses were used or if this study were repeated on a patient population with a lower BMI. The medications were well-tolerated and no participants reported stopping the medications; however, treatment adherence was not queried beyond self-report. Most relevant are the pre-treatment levels of LH and the associated T:LH ratio, which differ among the two groups. Furthermore, given the limited sample size of men on combination therapy, we are unable to perform higher level statistical analyses such as multivariate regression with a high degree of fidelity. Given the small sample size, results should be considered preliminary. There are many benefits to a multi-institutional approach, such as the one we employ here; however, with such a collaboration comes potential inter-facility measurement bias for lab testing and SAs. Even with a multi-institutional approach, our sample size remains relatively small—underscoring the innate challenges of studying this patient population. Future investigations with prospective study designs and a more robust number of participants may benefit from higher level statistical methodology that is not possible in this report, including propensity match score analysis."