PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	SEMEN QUALITY AND REPRODUCTIVE HORMONES IN FAROESE MEN – A CROSS-SECTIONAL POPULATION-BASED STUDY OF 481 MEN
AUTHORS	Petersen, Maria; Halling, Jonrit; Jørgensen, Niels; Jensen, Tina; Grandjean, Philiippe; Weihe, Pál

VERSION 1 - REVIEW

REVIEWER	Bernard Jégou, Inserm senior scientist/professor School of public health, IRSET-Inserm U1085, Rennes France
REVIEW RETURNED	18-Sep-2012

GENERAL COMMENTS	original and useful study completing the knowledges on the reproductive health status of the european citizen across the continent.

REVIEWER	Jim A. Mossman, Post-doctoral Research Associate, Brown University, USA. I declare no conflicting or competing interests.
REVIEW RETURNED	29-Nov-2012

THE STUDY	Some methods require clarifying, including reporting the results of statistical models in Supplementary Materials. The results section needs sub-heading for clarity. It jumps around a bit and is not systematically dealt with. The discussion section primarily reiterates the results and highlights methodological discrepancies between compared groups (which there were quite a few of). I suggest discussing the results in a wider context. I recommend reporting the statistical models in Supplementary materials so the reader can observe: (i) what confounding terms were included in the models, and (ii) what were the strengths of the relationships.
RESULTS & CONCLUSIONS	The results need to be better structured (use of sub-headings etc.) The discussion section is a little apologetic because the authors had to validate alternative methods/ assays between research cohorts, and they don't really focus on the overall significance of the work. Although, I do appreciate this is primarily a descriptive paper, however some of the authors' interesting findings were not

	discussed at all. I have highlighted these in my report.
GENERAL COMMENTS	This study describes the semen characteristics and reproductive hormone profiles of a population of Faroese men not previously selected on the basis of an (in)fertility phenotype. Comparisons are made to a larger cohort of Danish men and the authors conclude that semen characteristics are similar to Danish men, and typically low when compared to the World Health Organisation reference values. In addition, indices of reproductive hormone levels were lower than in a similar albeit larger Danish population. The authors speculate that genetic differences and environmental exposure may pose risk factors for semen characteristics in the Faroese population. I thought this descriptive paper was fairly straightforward and generally well written. I have a few comments that I feel require addressing before this paper is acceptable for publication. First, In the descriptive data in Table 1 there are some variables that are an order of magnitude different between F1 and F2, yet there appears to be no significant difference between these, which appears odd. For example, incidence of cryptorchidism treatment was 3.3 vs 7.4 and non-significant; hypospadias incidence was 0.4 vs 0.8 and again non-significant. I suggest double-checking these analyses to confirm the large differences are non-significant. Also comparison analyses between Faroese men and Danish men will be sensitive to very skewed sample sizes between groups, and this warrants discussion. For example, the incidence of hypospadias is two-fold greater in the collective Faroese men compared to Danes and this comparison is highly significant, possibly due to sample size differences, whereas between Faroese men, the comparison has the same magnitude of difference yet it is non-significant. It is interesting that the F1 vs F2 comparisons for fatherhood and experienced fertility problems are counter intuitive. Not only is the fatherhood significantly greater in the F1 cohort, but also this (F1) group had a higher incidence of fertility problems. Did the group with increased fertility problems also have increased fertility treatments? If so, the reported data are somewhat redundant and difficult to interpret. Given that the samples used to describe the population are not exactly the same as the samples used to describe the semen analysis data (e.g. 243 were questioned in F2 but only 240 provided a semen sample), it would be a good idea to restrict the semen analyses to only men who have both questionnaire and semen data for continuity. My main concern is about the measurement of sperm concentration in the two Faroese cohorts when a known error had occurred with the haemocytometer depth used. On reassessment of those samples suspected of being incorrect, I would suggest investigating the agreement between the values between the two measures (thawed vs original measure) and not just correlations. I assume this was done but no details are reported, only that they were accurate. I

	think a more obvious and visual way to confirm the correct samples were doubled for concentration would be to conduct a Bland and Altman plot (Bland and Altman, 1986 *Lancet* 1,307-310). This would include the data for the thawed samples of known correct calibration (0.1mm depth) and their comparable original concentration, plus the suspected 59 samples that were measured in undetermined haemocytometer chambers. The scatter plot will be the mean of the two concentration measures per individual on the x-axis against the difference of the two concentration measures on the y-axis. This plot will inform two critical things. First, whether the thawed and original sample concentrations agree and do not just correlate, and second, will reveal a sub-population of individuals that lie outside the range of 1.96* standard deviations of the mean difference of the known sample concentrations. E.g. plot mean vs difference for repeated concentrations of known chamber depths, calculate the overall mean difference and the standard deviation. Add horizontal lines +/- 1.96 SD of the mean difference. Then include the 59 suspect samples in the same way (original value and thawed value). If the original chamber depth was correct, the data should fall in the distribution of individuals within +/- 1.96 SD of the mean difference. If the chamber depth was incorrect, the respective data should fall outside this range, either above or below the 1.96 SD lines, and form a data sub-population (cluster of data, probably a straight line or arc). This sub-population will include all the individuals that require their original sperm concentration doubling. Inhibin B was also determined using two different kits (P8,L27). Was it the same kit but from different suppliers? This was a bit confusing. Are the between kit results comparable and has research validated their agreement? If so, this should be referenced. There were significant differences between the F1 and F2 cohorts for Inhibin B/FSH levels. Can the authors validate in any way by substantiating, perhaps in the Supplementary Information, that “inhibin/FSH results did not differ according to this”(P13,L21-22). Abstinence time was significantly different between the populations ((F1+F2) and D) and this is likely to impact on sperm number. Thus the differences between populations could be a consequence of abstinence time and not necessarily the differences between populations. However, in spite of a longer abstinence time in the F1+2 cohorts, they still produced similar numbers of sperm to the Danish population, suggesting sperm production rate is lower in the Faroese population compared to Danes. I feel this point should be included and discussed as it is important. Moreover, in spite of generally poorer semen characteristics in Faroese men compared to the Danish population, the fertility was in fact higher in Faroese than in Danes (P10, L41). I felt this should be discussed somewhat as it is an important finding in the study and is not mentioned in the discussion. Similarly, the criteria for ‘experienced fertility problems’ (Table 1) are

	different between the two groups and therefore comparisons between them are likely to be prone to error. The Faroese men, who described fertility problems on the basis of 6 months trying with unprotected intercourse, would be expected to have a higher incidence than Danes, who reported after trying for 1 year. Basically the logic follows that the men in the Danish population were more likely to have had a successful fertilization having tried for double the amount of time. Apologies if I have misinterpreted this from the text. This point could be discussed. The results of the analyses were all described in the text and I wonder if they may be described for clarity in a table, perhaps in the Supplementary Information. It would be informative for the reader to be able to observe the strengths of the relationships for semen characteristics reported in the results, especially with regards P12, L17-31. Plus, it would be easier to observe the terms in the models, since “significant confounders” (P9, L38) is a rather loose term. In addition, perhaps some sub-headings would help the results for clarity. They are not systematically dealt with at present and a little confusing. Page 13 lines 4-13 describe the effects of participant smoking habits and that of their mothers’. I know of at least one paper that has found similar non-significant effects of smoking on sperm characteristics (Povey *et al*., 2012 *Human Reproduction* 27, 2799-2806) that could be discussed along with the current results. Is this evidence of possible maternal in utero imprinting on the fetus? The discussion section at present seems to reiterate the results and offers little to put the paper in a wider perspective. I feel the authors should consider their findings in the broader discussion. Since semen ‘quality’ is a very subjective term and there is no consensus on what determines quality, I would suggest replacing the term with ‘characteristics’ throughout the manuscript. In the methods sperm motility was graded in three categories (progressive motile, local motile or motile: P7,L31). I think it would be a good idea to use the WHO categories (A,B,C and D) to show exactly what the motility scores are. Further, in the results section motile sperm (%) is reported. Is this the sum of progressively motile plus local motile, or the sum of WHO categories A plus B plus C, or other? Please specify in the Table legend or the accompanying methods. On P10, lines 10-20, these are results and appear in the methods section. I would consider entering this paragraph in the results section. The percentage of morphologically normal sperm was negatively associated with year of birth, suggesting younger men produced ejaculates with lower percentages of normal forms. This should be discussed (in the discussion) and not just the technical difficulties of teasing apart birth year from participant age, which is a technical issue. In the discussion, the authors mention an interesting finding

	that “semen quality has been reported to decline with increasing age” (P16,L24). However, the effects of participant age on sperm morphology followed the reverse pattern in the current study. Could the authors discuss this important point? I wonder whether the high autocorrelation between age and birth year could alone explain why controlling for birth year removed the trend for lower normal forms between Faroese cohorts. In other words, there was very little overlap between the cohorts for year of birth. I was slightly surprised that such a model would converge. The Table 1 descriptions of the sample populations reveal significant differences for almost all physical appearance, lifestyle, treatment factors etc. I think the authors should at least mention these major differences in the underlying biology between the populations of men that were compared for the study in the discussion. The comparisons were made to a ‘similar’ population, however the similarity is difficult to explain using the characteristics of the populations in Table 1 because they significantly differ in so many ways. I was a bit confused in the discussion (P17,L31) about in which cohort “men with suspected fertility problems may be overrepresented”. Please clarify whether this means all Faroese men or one of the Faroese cohorts. I thought one of the main take home messages of the paper, which would also place it into a broader perspective of declining semen characteristics over time, would be to mention that even within populations there can be marked differences in semen characteristics. It is perhaps unfortunate that different assessment methods/assays were employed between populations and even within cohorts which rendered the discussion a little apologetic in places. This however exemplifies one of the main problems when comparing semen characteristics globally and over time, namely that different methods are used spatially and temporally. I think this would add a strong discussion point and I advise the authors to consider this exact point. **Minor points:** In Table 3, what are the units for Free Testosterone? nmol/L? P17,L36. “Both high…testicular function” needs referencing. P18,L11. Be careful when speculating about motile sperm concentrations here. Earlier in the discussion the authors advise that such conclusions should be hesitated, suggesting these data are not reliable. Make sure PCBs have been defined before their acronym use. P7,L4. Criterion (singular). P8, L21. On the examination day.

	P8, L21. Centrifuged…..please provide *G and duration of centrifugation. P9,L16. …"declared for military service" remove "with". P10, L19. "as a confounder for sperm motility" P12, L29. "fathers group (p=0.6)" P12, L35. .."was detected in 15% of the F1 cohort". Needs subject. P13, L11. I was confused in this paragraph because it jumps around a bit. Does " Associations to….non-significant" refer to maternal smoking habits? Please specify. "morphology were non-significant…" "highly" is redundant.

REVIEWER	Jonatan Axelsson, PhD student, Lund University, Clinical Sciences, Malmö, Sweden No competing financial interests
REVIEW RETURNED	03-Dec-2012

THE STUDY	Study design: Yes and no. No to whether the semen quality would be lower due to POPs or genetic causes
RESULTS & CONCLUSIONS	Although the main message is clear, some results, not described thoroughly in the methods or aims makes the whole message somewhat less easy to catch. The article would benefit of excluding some of the extra analyses such as those mentioned in results that are not described earlier in the text.
REPORTING & ETHICS	How the study size was arrived at was not discussed. Did any included man have a missing value in any variable?
GENERAL COMMENTS	This article deals with semen quality and reproductive hormones in Faroese men compared with Danish men, all from the general population. General comments: It is an interesting study quite easy to read. However, the form of the article does not answer to whether the semen quality would be low due to POPs or genetic causes. Further, the analyses that gave some of the results were not discussed in the methods section. Specific comments: Page 3, line 27. In what ways is the study prospective? Page 4, line 13/14 - 17. I think the sentence aiming is somewhat wrong since

	the confirmation of low semen quality per se can be done without studying associations between semen quality and fertility, although association between the two latter makes the findings makes low semen quality important. Would suggest to consider two sentences. Reference number 9 does not seem to deal directly whith the association between semen quality and risk of infertility Page 5, line 27 (Table 1): How was the testis size calculated (sum of both or mean)? Page 7, line 31: Would change to progressively and locally motile Page 8, line 20-21. Would put a comma after day. Would change to "Sera were" or next line to "..it was analysed.." Page 9, line 34. Was motile spermatozoa = WHO:s a + b + c category = totally motile? Page 10, line 7 Why was not regression used for this specific variable? line 17-18. Was abstinence included as a confounder for all semen parameters? line 21. Is no influence = p>0.05 Page 11 line 5-6. Was overall motility progressive plus locally motile? Page 12, line 5. What was the reason to choose Mann-Whitney for this specific variable? In the methods section the authors describe this test was chosen for categorical variables which is seems not to be. line 6. was not abstinence included as a variable for motility? If not, why? line 9-15/16. This is not described in the methods section. It seems not to be totally appropiate to compare compare with WHO reference values since these were from proven fertile men, or at least this fact would possibly be mentioned. line 17. was this analysis adjusted for abstinence time? line 17-24. This seems not be mentioned in the methods section. line 22. Was it statistically significant? line 25-31. What is the specific reason to include these results. It does not seem to belong to the research question and is not fully described in the methods line 37-40. this does not seem to be part of the research question and the paper may benefit of excluding this part Page 13, line 19. Would put a comma after 2010 Page 15, line 13. Would change to SHBG .. was .. line 15-18/19. Would consider to remove this part line 25-38. Would consider to remove this part since it does not seem to be part of the main research question

	Page 16, line 11. Would be interesting with a dicussion with references about potential consequences for fertility line 19/20. Would put a comma after population line 34/35 Was this adjusted for abstinence time? line 36/37. Would put a comma after ours Page 17, line 10/11. Would write "variance of motility" Page 18, line 4- 9. Maybe it is hard to speculate whether POPs have caused a lower semen quality in other ways than the lower sperm motility in view of the current litterature. Even for motility other reasons may be present such as differences in life style, somewhat different ways of recruiting and age difference in spite of the danish refernce. Age intervals was almost separeted which somewhat may invalidate this adjustment.

VERSION 1 – AUTHOR RESPONSE

Response to comments from Dr. Jonathan Axelsson

The numbering of lines does not correspond to our word document of the manuscript. Due to this, we sometimes are in doubt what the reviewer refers to.

Comment: Is the overall study design appropriate and adequate to answer the research question? Yes and no. No to whether the semen quality would be lower due to POPs or genetic causes
Response: This article describes the semen quality of men from the Faroese population and the reasons for the low semen quality cannot be answered by the presented data. However, in the discussion section we speculate whether genetic and/or POPs could be contributing factors. This will be further explored in follow-up studies.

Comment: Although the main message is clear, some results, not described thoroughly in the methods or aims makes the whole message somewhat less easy to catch.
Response: Without knowing exactly which results Dr. Axelsson refers to, we assume the revision will clear the whole message.

Comment: The article would benefit of excluding some of the extra analyses such as those mentioned in results that are not described earlier in the text.
Response: We assume Dr. Axelsson refers to figure 2A and B and the text according to these Figures. Information that associations between year of birth and semen parameters were tested by linear regression has been added to the statistics section.

Comment: How the study size was arrived at was not discussed.
Response: Here we have to courteously disagree with the reviewer. In the method section this is described under these headings: Faroese men, examined 2007-09 (sub-group F1) and Faroese men, examined 2009-10 (sub-group F2).

Comment: Did any included man have a missing value in any variable?
Response: All men had semen variables assessed. However, morphology scoring was missing for 50 Faroese men and 20 Danes, which is reported in Table 2. Also, assessment of testis size was not performed on sub-group F2, which is stated in Table 1.

General comments:

Comment: It is an interesting study quite easy to read. However, the form of the article does not answer to whether the semen quality would be low due to POPs or genetic causes.
Response: As stated above this article describes the semen quality of men from the Faroese population and the reasons for the low semen quality cannot be answered by the presented data. However, in the discussion section we speculate whether genetic and/or POPs could be contributing factors. This will be further explored in follow-up studies.

Comment: Further, the analyses that gave some of the results were not discussed in the methods section.
Response: As stated above, we assume Dr. Axelsson refers to figure 2A and B and the text according to these Figures. Information that associations between year of birth and semen parameters were tested by linear regression has been added to the statistics section.

Specific comments:
Comment: Page 3, line 27. In what ways is the study prospective?
Response: We have clarified that it was a prospectively designed, cross-sectional study.

Comment: Page 4, line 13/14 - 17. I think the sentence aiming is somewhat wrong since the confirmation of low semen quality per se can be done without studying associations between semen quality and fertility, although association between the two latter makes the findings makes low semen quality important. Would suggest to consider two sentences.
Response: Here, we slightly disagree. The understanding of what is meant by low semen quality is mainly determined by studies of associations between semen quality and fertility chances, although the historical perspective of declining semen quality is also relevant. The studies showing geographical differences - referred to in the text – also concluded that low semen quality was prevalent because of comparison to fertility chances. Thus, we have kept our initial phrasings unchanged.

Comment: Reference number 9 does not seem to deal directly whith the association between semen quality and risk of infertility.
Response: We have replaced the original reference 9 with a more relevant.

Comment: Page 5, line 27 (Table 1): How was the testis size calculated (sum of both or mean)?
Response: The testis size was calculated as mean of two testes for each man. This information has been added to Table 1 and the text describing the physical examination.

Comment: Page 7, line 31: Would change to progressively and locally motile
Response: We have clarified that sperm were classified as progressive motile (WHO class AB motility), locally motile (WHO class C motility) or immotile (WHO class D motility).

Comment: Page 8, line 20-21. Would put a comma after day. Would change to "Sera were" or next line to "..it was analysed.."
Response: This has been done.

Comment: Page 9, line 34. Was motile spermatozoa = WHO:s a + b + c category = totally motile?
Response: Yes, it was AB+C, which has been clarified both in the statistics section and in the footnote of Table 2.

Comment: Page 10, line 7. Why was not regression used for this specific variable?
Response: We assume Dr. Axelsson refers to the comparison to the distribution of the WHO reference data. We did not have the exact dataset on which the WHO reference levels were based but only the distribution. Thus, the distribution differences between the Faroese men and the WHO

reference was tested by ci square analyses, which is stated in the statistics section.

Comment: line 17-18. Was abstinence included as a confounder for all semen parameters?
Response: The abstinence time was included as a confounder for sperm concentration, semen volume and total sperm count; duration from ejaculatation to assessesment was included when evaluating motility; no confounders were included when morphology was evaluated. This has been clarified in the statistics section.

Comment: line 21. Is no influence = p>0.05
Response: p >0.05 was considered as no influence. The information has been added to the statistics section.

Comment: Page 11, line 5-6. Was overall motility progressive plus locally motile?
Response: Yes, it was AB+C, which has been clarified both in the statistics section and in the footnote of Table 2.

Comment: Page 12, line 5. What was the reason to choose Mann-Whitney for this specific variable? In the methods section the authors describe this test was chosen for categorical variables which is seems not to be.
Response: We used the Mann- Whitney test because adjusting for confounders was not needed. However, we repeated the analysis using linear regression, and the result is now based on this.

Comment: line 6. Was not abstinence included as a variable for motility? If not, why?
Response: We assume you refer to Table 1. Motility was adjusted for duration from time of ejaculation to assessment of motility. Motility was not affected by duration of ejaculation abstinence (which was also the case for the Danish/Nordic studies cited as references 4, 5, 12, 17, and 21).

Comment: line 9-15/16. This is not described in the methods section. It seems not to be totally appropriate to compare with WHO reference values since these were from proven fertile men, or at least this fact would possibly be mentioned.
Response: In the methods section we stated that the Faroese sperm count distribution was compared with the distribution for fertile men reported as the WHO reference group. We have kept this information. Furthermore, the findings are interpreted in the discussion section together with comparison between Danes and Faroese men. We still feel that it is relevant to compare the data from our population studies with the WHO reference population.

Comment: line 17. Was this analysis adjusted for abstinence time?
Response: No this analysis was not adjusted for abstinence time. Please, also see our previous answer regarding adjustment for ejaculation abstinence time.

Comment: line 17-24. This seems not be mentioned in the methods section.
Response: Please, see our previous response regarding figure 2A and B. This has now been the statistics section.

Comment: line 22. Was it statistically significant
Response: The confounder adjusted estimates indicated higher motility in men being fathers in comparison to non-fathers (p<0.0005), which is described in the results section “Fertility”.

Comment: line 25-31. What is the specific reason to include these results. It does not seem to belong to the research question and is not fully described in the methods line 37-40. this does not seem to be part of the research question and the paper may benefit of excluding this part
Response: The physical examination can give clues to whether any specific variable potentially

affecting semen quality is high in the Faroese population studied or not. Furthermore, in our opinion such information is part of the description of the study population, and thus we regret that a physical examination was not performed in both cohorts. Thus, we hope you accept that we have kept the information unchanged.

Comment: Page 13, line 19. Would put a comma after 2010
Response: This has been done.

Comment: Page 15, line 13. Would change to SHBG .. was ..
Response: This has been done.

Comment: line 15-18/19. Would consider to remove this part line 25-38. Would consider to remove this part since it does not seem to be part of the main research question
Response: The hormone levels can reflect the semen quality and since there was significant difference in many of the hormones, we feel that this issue has to be addressed. Furthermore, reproductive hormone levels also reflect testicular function besides semen quality, and we have changed the title of the manuscript accordingly.

Comment: Page 16, line 11. Would be interesting with a dicussion with references about potential consequences for fertility
Response: We are in doubt what Dr. Axelsson refers to. Our manuscript is mainly descring semen quality and reproductive hormone level. However, in the interpretation in the discussion section we have expanded this to include additional information about the number of Faroese men that had experienced fertility problems. Furthermore, we have kept the comparison to the WHO reference population of fertile men.

Comment: line 19/20. Would put a comma after population
Response: This has been done.

Comment: line 34/35 Was this adjusted for abstinence time?
Response: No this analysis was not adjusted for abstinence time. Please, also see our previous answer regarding adjustment for ejaculation abstinence time.

Comment: line 36/37. Would put a comma after ours
Response: This has been done.

Comment: Page 17, line 10/11. Would write "variance of motility"
Response: This has been done.

Comment: Page 18, line 4- 9. Maybe it is hard to speculate whether POPs have caused a lower semen quality in other ways than the lower sperm motility in view of the current litterature. Even for motility other reasons may be present such as differences in life style, somewhat different ways of recruiting and age difference in spite of the danish refernce. Age intervals was almost separeted which somewhat may invalidate this adjustment.
Response: As we state in the discussion, the reason for the low testicular function of the Faroese men is currently unknown. However, we find it relevant to point to POPs as a potential contributing factor because the Faroese population is highly exposed to POPs derived from traditional marine food.

Response to comments from Jim A. Mossman

Comment: The results section needs sub-heading for clarity. It jumps around a bit and is not systematically dealt with.
Response: Sub-headings have been added.

Comment: The discussion section primarily reiterates the results and highlights methodological discrepancies between compared groups (which there were quite a few of). I suggest discussing the results in a wider context.
Response: We agree that technical aspects are discussed thoroughly which we find relevant when reporting results from population studies to allow the reader to understand strengths and weaknesses. However, we have also included perspectives of our findings: semen quality at a low lever even compared to Danes that hitherto have been shown to have particularly poor semen quality; semen quality at a low level among the Faroese men when compared with the WHO fertile reference population, and we have also suggested potential explanatory factors that should/could be investigated in following studies.

Comment: I recommend reporting the statistical models in Supplementary materials so the reader can observe: (i) what confounding terms were included in the models, and (ii) what were the strengths of the relationships.
Response: In the statistics section of the manuscript we have added specific information that ejaculation abstinence time was included as a confounder for sperm concentration, semen volume and total sperm count; duration from ejaculation to assessment was included when evaluating motility; no confounders were included when morphology was evaluated. Also, information of which other confounders were tested can be seen in the statistics section. Whether the reporting of statistical models should be in Supplementary information rather than in the main text may be an editorial decision, which we leave for the editor to decide. However, in the revised version we have kept the information in the main text similarly to what we previously have done (e.g. ”Jørgensen et al, BMJ Open 2012: Human semen quality in the new millenium”)

Comment: The results need to be better structured (use of sub-headings etc.)
Response: Sub-headings have been added.

Comment: The discussion section is a little apologetic because the authors had to validate alternative methods/ assays between research cohorts, and they don't really focus on the overall significance of the work. Although, I do appreciate this is primarily a descriptive paper, however some of the authors' interesting findings were not discussed at all. I have highlighted these in my report.
Response: Please, see our second response to Dr. Axelsson above.

FULL REVIEW
Comment: This study describes the semen characteristics and reproductive hormone profiles of a

population of Faroese men not previously selected on the basis of an (in)fertility phenotype. Comparisons are made to a larger cohort of Danish men and the authors conclude that semen characteristics are similar to Danish men, and typically low when compared to the World Health Organisation reference values. In addition, indices of reproductive hormone levels were lower than in a similar albeit larger Danish population. The authors speculate that genetic differences and environmental exposure may pose risk factors for semen characteristics in the Faroese population. I thought this descriptive paper was fairly straightforward and generally well written. I have a few comments that I feel require addressing before this paper is acceptable for publication.
Response: Our detailed responses can be seen above and below as a point-by-point response to the reviewers' comments.

Comment: First, In the descriptive data in Table 1 there are some variables that are an order of magnitude different between F1 and F2, yet there appears to be no significant difference between these, which appears odd. For example, incidence of cryptorchidism treatment was 3.3 vs 7.4 and non-significant; hypospadias incidence was 0.4 vs 0.8 and again non-significant. I suggest double-checking these analyses to confirm the large differences are non-significant.
Response: We have double-checked these analyses and the other tables as well. The p-values Dr. Mossman mentioned were in fact wrong because the answer "don't know" was not set as a missing variable.

Comment: Also comparison analyses between Faroese men and Danish men will be sensitive to very skewed sample sizes between groups, and this warrants discussion. For example, the incidence of hypospadias is two-fold greater in the collective Faroese men compared to Danes and this comparison is highly significant, possibly due to sample size differences, whereas between Faroese men, the comparison has the same magnitude of difference yet it is non-significant.
Response: The sample size of the Faroese group was the largest we could include. However, compared to the Danish population the frequencies reported in e.g. Table 1 may be less accurate. We have included this aspect in the discussion section.

Comment: It is interesting that the F1 vs F2 comparisons for fatherhood and experienced fertility problems are counter intuitive. Not only is the fatherhood significantly greater in the F1 cohort, but also this (F1) group had a higher incidence of fertility problems. Did the group with increased fertility problems also have increased fertility treatments? If so, the reported data are somewhat redundant and difficult to interpret.
Response: We do not have information on fertility problems and fertility treatment. The only question that we have asked giving us this information is: Have you ever had regular intercourse without use of contraception for at least 6 months. The age difference can be one explanation to the difference between these two groups in regard to fatherhood.

Comment: Given that the samples used to describe the population are not exactly the same as the samples used to describe the semen analysis data (e.g. 243 were questioned in F2 but only 240 provided a semen sample), it would be a good idea to restrict the semen analyses to only men who have both questionnaire and semen data for continuity.
Response: We find your comment highly relevant and now we have re-done all the analyses with 240 men (F2).

Comment: My main concern is about the measurement of sperm concentration in the two Faroese cohorts when a known error had occurred with the haemocytometer depth used. On reassessment of those samples suspected of being incorrect, I would suggest investigating the agreement between the values between the two measures (thawed vs original measure) and not just correlations. I assume this was done but no details are reported, only that they were accurate. I think a more obvious and visual way to confirm the correct samples were doubled for concentration would be to conduct a

Bland and Altman plot (Bland and Altman, 1986 Lancet 1,307-310). This would include the data for the thawed samples of known correct calibration (0.1mm depth) and their comparable original concentration, plus the suspected 59 samples that were measured in undetermined haemocytometer chambers.
The scatter plot will be the mean of the two concentration measures per individual on the x-axis against the difference of the two concentration measures on the y-axis. This plot will inform two critical things. First, whether the thawed and original sample concentrations agree and do not just correlate, and second, will reveal a sub-population of individuals that lie outside the range of 1.96* standard deviations of the mean difference of the known sample concentrations. E.g. plot mean vs difference for repeated concentrations of known chamber depths, calculate the overall mean difference and the standard deviation. Add horizontal lines +/- 1.96 SD of the mean difference. Then include the 59 suspect samples in the same way (original value and thawed value). If the original chamber depth was correct, the data should fall in the distribution of individuals within +/- 1.96 SD of the mean difference. If the chamber depth was incorrect, the respective data should fall outside this range, either above or below the 1.96 SD lines, and form a data sub-population (cluster of data, probably a straight line or arc). This sub-population will include all the individuals that require their original sperm concentration doubling.
Response: A detailed analysis was done but not included in the manuscript to make it simple. We have kept the information in the manuscript unchanged, but have added detailed information as Supplementary Materials, in which we have presented the original counting's, the recounting and the relative difference between these, and then finally the conclusion of whether to adjust the original result or not. Please, see the Supplementary information for details.

Comment: Inhibin B was also determined using two different kits (P8,L27). Was it the same kit but from different suppliers? This was a bit confusing. Are the between kit results comparable and has research validated their agreement? If so, this should be referenced. There were significant differences between the F1 and F2 cohorts for Inhibin B/FSH levels. Can the authors validate in any way by substantiating, perhaps in the Supplementary Information, that "inhibin/FSH results did not differ according to this" (P13,L21-22).
Response: In the methods section we have expanded the information regarding Inhibin B assessments, including information that internal validation did not indicate a need to correct for inter-assay variation, including no need for correction between the kit material from the two suppliers. The difference in Inhibin B levels between the two cohorts were non-significant, and difference in Inhibin B/FSH-ratio is mainly driven by differences in FSH levels. Furthermore, the non-significantly higher Inhibin B in F2 was accompanied by lower FSH and vice versa for F1, as expected according to the endocrine negative feed-back loop.

Comment: Abstinence time was significantly different between the populations ((F1+F2) and D) and this is likely to impact on sperm number. Thus the differences between populations could be a consequence of abstinence time and not necessarily the differences between populations. I feel this point should be included and discussed as it is important.
Response: Duration of abstinence affects semen volume, sperm concentration and total sperm counts, and needs to be taken into account. Therefore, the comparisons between groups were adjusted for abstinence in the regressions analyses. We have described this in the statistics section. Please, also see our response to Dr. Axelsson regarding abstinence time. Further taking into account that the duration of abstinence was longer in the Faroese men the unadjusted sperm concentration should be higher in these men if the groups were not different, but the opposite was the case. Despite of higher semen volume in the Faroese men and a longer abstinence time they still only produced similar total sperm counts as the Danish population, suggesting that sperm production rate is lower in the Faroese population compared to Danes.

Comment: Moreover, in spite of generally poorer semen characteristics in Faroese men compared to

the Danish population, the fertility was in fact higher in Faroese than in Danes (P10, L41). I felt this should be discussed somewhat as it is an important finding in the study and is not mentioned in the discussion.
Response: As shown in table 1 more Faroese men answered yes to the question “have you ever caused a pregnancy”, but the men (F1) were 25 years old contrary to the Danish men being 19 years old, which may be the major explanation for the difference between the groups. Further, the Faroese people are in general younger when parenting for the first time. Additionally, more Faroese men reported having experienced fertility problems than the Danish men. This could also be a consequence of the higher age, but it would also be agreement with the lower sperm concentration found in the Faroese men, but it should be taken into account that the Faroese men were asked about fertility problems on the based on 6 months trying with unprotected intercourse and for the Danes it was based on 1 year trying with unprotected intercourse. These aspects have been added to the discussion section.

Comment: Similarly, the criteria for ‘experienced fertility problems’ (Table 1) are different between the two groups and therefore comparisons between them are likely to be prone to error. The Faroese men, who described fertility problems on the basis of 6 months trying with unprotected intercourse, would be expected to have a higher incidence than Danes, who reported after trying for 1 year. Basically the logic follows that the men in the Danish population were more likely to have had a successful fertilization having tried for double the amount of time. Apologies if I have misinterpreted this from the text. This point could be discussed.
Response: Please, see our previous response.

Comment: The results of the analyses were all described in the text and I wonder if they may be described for clarity in a table, perhaps in the Supplementary Information. It would be informative for the reader to be able to observe the strengths of the relationships for semen characteristics reported in the results, especially with regards P12, L17-31. Plus, it would be easier to observe the terms in the models, since “significant confounders” (P9, L38) is a rather loose term. In addition, perhaps some sub-headings would help the results for clarity. They are not systematically dealt with at present and a little confusing.
Response: Age, height, weight, BMI and for F1 also testis size are shown in Table 1. Sub-headings have been added.

Comment: Page 13 lines 4-13 describe the effects of participant smoking habits and that of their mothers’. I know of at least one paper that has found similar non-significant effects of smoking on sperm characteristics (Povey et al., 2012 Human Reproduction 27, 2799-2806) that could be discussed along with the current results. Is this evidence of possible maternal in uterus imprinting on the fetus? The discussion section at present seems to reiterate the results and offers little to put the paper in a wider perspective. I feel the authors should consider their findings in the broader discussion.
Response: In the discussion we have added this reference and other from studies of young men from the general populations in various countries that showed the same. Since our finding belongs to what can be considered “expected”, we only added few comments to the discussion about this.

Comment: Since semen ‘quality’ is a very subjective term and there is no consensus on what determines quality, I would suggest replacing the term with ‘characteristics’ throughout the manuscript.
Response: In similar studies describing ”semen quality” this term has been used, whereas the term ”characteristics” is rarely used. Furthermore, in our manuscript semen quality refers back to the semen variables making it clear what is actually meant. Thus, we have kept the term “semen quality” unchanged.

Comment: In the methods sperm motility was graded in three categories (progressive motile, local motile or motile: P7,L31). I think it would be a good idea to use the WHO categories (A,B,C and D) to show exactly what the motility scores are.
Response: In accordance with the WHO guideline from 2010, the laboratory in Copenhagen has for several years (also before 2010) only categorized progressive motility in one category. As our laboratory followed the instructions from the Danish laboratory in group F2 only 3 categories were used and thus we do not have the possibility to report 4 categories.

Comment: Further, in the results section motile sperm (%) is reported. Is this the sum of progressively motile plus local motile, or the sum of WHO categories A plus B plus C, or other? Please specify in the Table legend or the accompanying methods.
Response: We have clarified that sperm were classified as progressive motile (WHO class AB motility), locally motile (WHO class C motility) or immotile (WHO class D motility) both in the methods section and in Table 2.

Comment: On P10, lines 10-20, these are results and appear in the methods section. I would consider entering this paragraph in the results section.
Response: Even though we understand your point, we do feel that this section is more appropriate here. When looking at the effect of abstinence time we found that the effect was not the same throughout the period and thus the abstinence time was entered as piecewise linear functions. If we do not have these “results” in connection with this, we feel that the reader will have difficulties to see what has been done and why. Furthermore, effect of abstinence is not the main focus of this manuscript.

Comment: The percentage of morphologically normal sperm was negatively associated with year of birth, suggesting younger men produced ejaculates with lower percentages of normal forms. This should be discussed (in the discussion) and not just the technical difficulties of teasing apart birth year from participant age, which is a technical issue.
Response: We have added the information that the finding is compatible with a prenatal adverse exogenous effect on the developing testis where the adverse exposures have become more prevalent.

Comment: In the discussion, the authors mention an interesting finding that “semen quality has been reported to decline with increasing age” (P16,L24). However, the effects of participant age on sperm morphology followed the reverse pattern in the current study. Could the authors discuss this important point?
Response: In our study we did not find an effect of age on the morphology, only an effect of the year of birth. However, we used the sentence “semen quality has been reported to decline with increasing age” to start a discussion whether the age difference between the Faroese and Danish population could have caused the detected difference between the two populations, and concluded that this is less likely because the age difference was quite small.

Comment: I wonder whether the high autocorrelation between age and birth year could alone explain why controlling for birth year removed the trend for lower normal forms between Faroese cohorts. In other words, there was very little overlap between the cohorts for year of birth. I was slightly surprised that such a model would converge.
Response: We are in doubt what to add to the manuscript. We reported the findings when correction for birth year and age, respectively, was done, and the effect of birth year remained strongest as stated in the results section.

Comment: The Table 1 descriptions of the sample populations reveal significant differences for almost all physical appearance, lifestyle, treatment factors etc. I think the authors should at least mention

these major differences in the underlying biology between the populations of men that were compared for the study in the discussion. The comparisons were made to a 'similar' population, however the similarity is difficult to explain using the characteristics of the populations in Table 1 because they significantly differ in so many ways.
Response: It is correct that many of the differences were statistically significant: E.g. $p<0.0005$ for height, however, reflecting a difference in median of 1.6 cm, which from a biological point of view can be regarded as similar. Furthermore, the tested the effect of these potential confounders as described in the statistics sections. We have added comments about this in the discussion section.

Comment: I was a bit confused in the discussion (P17,L31) about in which cohort "men with suspected fertility problems may be overrepresented". Please clarify whether this means all Faroese men or one of the Faroese cohorts.
Response: The F1 were randomly selected and therefore we cannot exclude that men with suspicion of infertility may have been more interested in participating in the study. Therefore, we also described results based on a subgroup of men without any prior knowledge of andrologically diseases, including suspected fertility problems to control for this potential problem. The discussion section has been expanded regarding this aspect.

Comment: I thought one of the main take home messages of the paper, which would also place it into a broader perspective of declining semen characteristics over time, would be to mention that even within populations there can be marked differences in semen characteristics. It is perhaps unfortunate that different assessment methods/assays were employed between populations and even within cohorts which rendered the discussion a little apologetic in places. This however exemplifies one of the main problems when comparing semen characteristics globally and over time, namely that different methods are used spatially and temporally. I think this would add a strong discussion point and I advise the authors to consider this exact point.
Response: We have added a paragraph in the discussion about this point.

Minor points:
Comment: In Table 3, what are the units for Free Testosterone? nmol/L?
Response: Yes. This has been added to the table.

Comment: P17,L36. "Both high…testicular function" needs referencing.
Response: References have been added.

Comment: P18,L11. Be careful when speculating about motile sperm concentrations here. Earlier in the discussion the authors advise that such conclusions should be hesitated, suggesting these data are not reliable.
Response: Other studies have mainly seen affect on motility and we showed what we found. However, we repeated our reservations.

Comment: Make sure PCBs have been defined before their acronym use.
Response: This has been done now.

Comment: P7,L4. Criterion (singular).
Response: This has been corrected.

Comment: P8, L21. On the examination day.
Response: This has been done.

Comment: P8, L21. Centrifuged…..please provide *G and duration of centrifugation.
Response: 3000 g, 10 minutes. This information has been added to the manuscript.

Comment: P9,L16. …"declared for military service" remove "with".
Response: This has been done.

Comment: P10, L19. "as a confounder for sperm motility"
Response: This has been done.

Comment: P12, L29. "fathers group (p=0.6)"
Response: This has been done.

Comment: P12, L35. .."was detected in 15% of the F1 cohort". Needs subject.
Response: This has been done.

Comment: P13, L11. I was confused in this paragraph because it jumps around a bit. Does " Associations to….non-significant" refer to maternal smoking habits? Please specify.
"morphology were non-significant…" "highly" is redundant.
Response: We have added the information "between mother's smoking" to make sure that it is clear that this is the mothers case here.

VERSION 2 – REVIEW

REVIEWER	Jim A. Mossman, Post-doctoral researcher Brown University, Rhode Island, USA I declare no competing interests.
REVIEW RETURNED	31-Jan-2013

THE STUDY	I have outlined that the results of the analyses should be included in a table, or tables in the Supplementary Materials. This will make the analyses easier to follow and the reader will be able to observe the strengths of the reported relationships. Plus, the authors make a number of statements about what were confounders and what were included in the statistical models. This could be easily facilitated if the model results are described in tables.
RESULTS & CONCLUSIONS	Supplementary Materials Tables would greatly facilitate the description and reader understanding of the data.
GENERAL COMMENTS	General: I am generally happy with the changes made to the manuscript, and am pleased that sub-headings have been introduced for ease of reading. I am slightly disappointed that the results of the analyses have not been presented in tables in the Supplementary Materials. This would make the analyses completely transparent, plus the reader would be able to observe the strengths of the relationships. I will leave this to the discretion of the Editor. I am glad the authors have toned-down the possible associations between genetic and environmental exposure as risk factors for semen characteristics, since Dr Axelsson is correct to highlight that the paper does not directly address this issue. The discussion is much better in the revised form. Minor:

	P10, L24: do you mean a motility measure was missing for one man? P13, L6: needs punctuation. Eg.“There were no significant associations between maternal smoking, sperm motility, and sperm morphology.” If you use the existing sentence “highly non-significant”, “highly” is redundant P13,L28: “…did not differ according to this is described in the methods section” P14,L55: “was considerably” space required. P16,L14: a closed parenthesis is required. P16,L7-34: “per women” need to be changed to “per woman” throughout. P17,L13: replace “almost similar protocols” with “similar protocols” “almost” is redundant.

VERSION 2 – AUTHOR RESPONSE

P10, L24: do you mean a motility measure was missing for one man? "Yes and this has now been changed"
P13, L6: needs punctuation. Eg.“There were no significant associations between maternal smoking, sperm motility, and sperm morphology.”
If you use the existing sentence “highly non-significant”, “highly” is redundant "We changed the sentence according to the reviewers suggestion"
P13,L28: “…did not differ according to this is described in the methods section” "This has been changed"

P14,L55: “was considerably” space required. "This has been changed"
P16,L14: a closed parenthesis is required. "There was not supposed to be any paranthesis so it has beed deleted"
P16,L7-34: “per women” need to be changed to “per woman” throughout. "This has been changed"
P17,L13: replace “almost similar protocols” with “similar protocols” “almost” is redundant. "We agree and have deleted "almost""