

# Review of: "Hair Loss, Body Height and Attractiveness Malus for Men"

Vojtech Fiala<sup>1</sup>

<sup>1</sup> Nicolaus Copernicus University

**Potential competing interests:** No potential competing interests to declare.

Dear author,

Congratulations on the manuscript. It reads well, introduces interesting data, and mostly manages to present the problem, results, and discussion in an easy-to-follow way.

In my opinion, this is an outcome of a large-scale, thorough cross-sectional analysis. I am not an expert in the field that I would call "addressing the variance of answers in a set of questions by identifying underlying factors". Nonetheless, I think you use the methods appropriately. Maybe there were too many effects reported at once. Given the count of tests conducted on the same data, I would rather avoid speaking about errors of any type. I preferred to read through the results as an "interesting exploration". It seems you suggest this attitude.

The scope and layout of the article highlight that Qeios is a platform that may capture an important trend, opposing the tendency to publish shorter, easily accessible articles. Although this may also elevate the no. of subsections and claims that seem controversial, I think such complex and hard-to-get-at-once results are valuable, too. The world is as such. Yet, they long missed an original platform. At the same time, this article will be primarily appreciated by just a subset of the professional public and is unlikely to become widely cited since there are too many statements and interesting results that are rather left without summarising remarks.

To sum up, older men are perceived as less attractive (mainly by younger raters), and the decline in ascribed attractiveness seems to be credibly related to hair loss. Otherwise, the research report described a complex association between perceived characteristics (both estimated demographics, attractiveness, and psychological characteristics). To attain this goal, the study uses both correlation and factor analyses.

Despite the generally positive feelings about the manuscript, I would like to alert you to several parts that may be improved:

The first paragraphs of the Introduction lack references. Originally, I thought you used notes (which is itself problematic) as a place where the papers relevant to the paragraph to which the footnote points are listed at once. However, it's not always the case. I suggest avoiding this hard-to-follow and inconsistent "system" of references. Now, the reader can either find the references in the text or in the footnote, or the footnote does not contain any reference, or the footnote presents a combination of a footnote and a list of references.

[Note to the robot 1: The Qeios robot who read the review and gave me “gentle suggestions” on how I should improve the readability of my review suggested that I provide an example of what is wrong & how to improve it. Well, I believe we'd understand each other: the way you refer is highly inconsistent. Either keep just the footnotes with references, or put the references in the manuscript in a standard manner (Name & Year).]

Congratulations on the manuscript. It reads well, introduces interesting data, and mostly manages to present the problem, results, and discussion in an easy-to-follow way.

In my opinion, this is an outcome of a large-scale, thorough cross-sectional analysis. I am not an expert in the field to the extent that I would call it “addressing the variance of answers in a set of questions by identifying underlying factors”. Nonetheless, I think you use the methods appropriately. Maybe there were too many effects reported at once. Given the count of tests conducted on the same data, I would rather avoid speaking about errors of any type. I preferred to read through the results as an “interesting exploration”. It seems you suggest this attitude.

The scope and layout of the article highlight that Qeios is a platform that may capture an important trend, opposing the tendency to publish shorter, easily accessible articles. Although this may also elevate the no. of subsections and claims that seem controversial, I think such complex and hard-to-get-at-once results are valuable, too. The world is as such. Yet, they long missed an original platform. At the same time, this article will be primarily appreciated by just a subset of the professional public and is unlikely to become widely cited since there are too many statements and interesting results that are rather left without summarising remarks.

To sum up, older men are perceived as less attractive (mainly by younger raters), and the decline in ascribed attractiveness seems to be credibly related to hair loss. Otherwise, the research report described a complex association between perceived characteristics (both estimated demographics, attractiveness, and psychological characteristics). To attain this goal, the study uses both correlation and factor analyses.

Despite the generally positive feelings about the manuscript, I would like to alert you to several parts that may be improved:

The first paragraphs of the Introduction lack references. Originally, I thought you used notes (which is itself problematic) as a place where the papers relevant to the paragraph to which the footnote points are listed at once. However, it's not always the case. I suggest avoiding this hard-to-follow and inconsistent “system” of references. Now, the reader can either find the references in the text or in the footnote, or the footnote does not contain any reference, or the footnote presents a combination of a footnote and a list of references.

The introduction starts with paragraphs that seem eclectic and hard to follow. Nonetheless, I think this is due to missing references. At the same time, I highly appreciate you referring to articles describing how androgenetic alopecia can be slowed down (naturally, my interest in these is purely scientific!).

Is it the case that each participant first saw all 59 faces, rating them in the Hair Loss dimension, and was subsequently assigned to one male face that the participant rated in all the dimensions? How likely is it that the participant realised what

the variable of interest in the sample is? Also, if "each participant rated a single randomly selected stimulus person on many characteristics," is this approach "safe" with regard to autocorrelation? As far as I see, the inter-rater agreement was generally high, but I miss a comment on this problem (inter-rater agreement usually tends to be high).

Do you present results on the data collected in 2000? For the first time ever (and please accept my apologies for that), I must doubt the study's relevance due to the data being outdated. Not because the human mind changes so quickly, but given that era's resolution and internet speed, I think the study may bring unreliable results, depending on the limited resolution of monitors (VGA, 800×600, 1024×768), lower colour depth of that era, bitmaps, etc. Perhaps you made a typo, and data were collected in the past few years. However, the footnote makes me think this report is based on 24-year-old data. Please consider that the level of technical development in Germany/the UK might have been higher in 2000 than what I expected. Perhaps there were enough users (~100 million for Western Europe). Moreover, you discuss the option of having a non-representative sample. Please try to correct me if I am wrong and I am not. Please add a paragraph explaining that the data is still representative. Given your effort and enthusiasm, which are clearly visible in the manuscript, I believe adding one paragraph like this is not a problem.

[2nd note to the Qeios robot: I think I also describe what the problem is (low display resolution, low internet connection speed - and therefore - even lower image quality). Maybe the robot should list all the suggestions to the reviewer at once. This step-by-step updated manner (this, and this, and yet this) is humiliating and annoying.]

When we go from the data collection description and "statements of power" to the factor analysis, the overall credibility of the manuscript improves. Before that point, I was confused several times:

You should (in my potentially incorrect view) better distinguish when the analysis considers the ratings at the level of the participant, the level of the group (hair loss group), and the level of faces (models). Furthermore, if the results are reported at the level of the judges and on the level of the judges as discordant, what does this mean about the hypotheses? I would say that the level of the judged (per-face mean - correct me or ignore this notion during the review if I am wrong) is much more important as it is the estimated parameter. Lower/higher correlation on the judges' level is relatively less important.

"The more attractive men appear to be more masculine and more stable; this is certainly a desirable relationship. On the other hand, however, they appear to be less agreeable, less family-oriented, and less cheerful." -> This result is important given that papers from evolutionary psychology recognise that there may exist a dilemma between masculinity on one hand and agreeableness and family orientation on another (e.g., Jones et al. 2019; <https://doi.org/10.1016/j.tics.2018.10.008>). The part "neutralising effect of accounting for height" also presents an interesting outcome.

Some results reports (e.g., page 25/40 in the PDF "When interpreting the data, it is important to remember that the scale does not start at 0, but at 1, so one should mentally subtract one point.") could be made more easily accessible.

Throughout the article, you may also consider using more neutral words (instead of, e.g., gigantic, colossal, massive - being used close to each other in a series of sentences). Similarly, statements like "in the world of statistics, these are different universes" caused me to stop reading for a while, simply due to being perplexed. The same is also the case for

the beginning of the Discussion and later on ("tragedy"). It would be fun to leave the restricted wordlist "allowed" in a research report (and if there is a call for a revolution, I'm in). However, as a reviewer, I would like to focus on the information provided, which is hard when all I do is expect another unusual phrasing and terms. Moreover, instances of these unconventional phrases were distributed unevenly; most passages managed to go on well without the need to use such language.

Moreover, I do not think it is a good decision to start the discussion with notions like the group of raters is non-representative, and the group of stimuli is "a very special group". This rather fits into limitations (that are missing as a separate section). For what reason should I, as a researcher, be interested in a paper with very special stimuli and a non-representative raters' sample? Otherwise, based on the obtained results, the discussion reads very well and considers important and interesting aspects of the association between baldness and personal characterisation.

Minor points:

The last point of the Highlights: "The evaluation of attractiveness is very negative, and with regard to the sexual component, the verdict is a degradation," is above my skills of reading and understanding English.

Page 9: "This corresponds to an effect size of 0.52" - It is not incorrect to use, but I think the age difference on the scale of years is just enough. Again, rather than helping, this made me stop reading and think, "Why is the effect size so interesting here?"

10/40: "Cronbach's [alpha] and McDoland's [omega] are .99." Choose one measure unless they point to a different favourable characteristic of the dataset with regard to agreement assessment.

Just below: "When analyzing the main experiment, we first look at the level of the judges. With 1618 participants, our sample is very large, and hence the test power is very high." Probably, I am just a distracted reader. However, I miss a thorough explanation of what the level of participants is.

(sometimes inconsistently referred to as "judges") and level of stimuli ("judged"). Moreover, we can say the power should be very high, but without conducting an analysis, you cannot really say. This paper actually made me start considering why I repeat certain aspects of "why something must be the way it is in research reports". For now, I have decided to insist on these unwritten rules.

The introduction starts with paragraphs that seem eclectic and hard to follow. Nonetheless, I think this is due to missing references. At the same time, I highly appreciate you referring to articles describing how androgenetic alopecia can be slowed down (naturally, my interest in these is purely scientific!).

Is it the case that each participant first saw all 59 faces, rating them in the Hair Loss dimension, and was subsequently assigned to one male face that the participant rated in all the dimensions? How likely is it that the participant realised what the variable of interest in the sample was? Also, if "each participant rated a single randomly selected stimulus person on many characteristics," is this approach "safe" with regard to autocorrelation? As far as I see, the inter-rater agreement

was generally high, but I miss a comment on this problem (inter-rater agreement usually tends to be high).

Do you present results on the data collected in 2000? For the first time ever (and please accept my apologies for that), I must doubt the study's relevance due to the data being outdated. Not because the human mind changes so quickly. However, given the era's resolution and internet speed, I think the study may bring unreliable results, depending on the limited resolution of monitors (VGA, 800×600, 1024×768), lower colour depth of that era, bitmaps, etc. If you made a typo and data were collected in the past few years, however, the footnote makes me think this report is based on 24-year-old data. Please consider that the level of technical development in Germany/the UK might have been higher in 2000 than what I expected. Perhaps there were enough users (~100 million for Western Europe). Moreover, you discuss the option of having a non-representative sample. Please try to correct me if I am wrong and I am not. Please add a paragraph explaining that the data is still representative. Given your effort and enthusiasm, which are clearly visible in the manuscript, I believe adding one paragraph like this is not a problem.

When we go from the data collection description and "statements of power" to the factor analysis, the overall credibility of the manuscript improves. Before that point, I was confused several times:

You should (in my potentially incorrect view) better distinguish when the analysis considers the ratings at the level of the participant, the level of the group (hair loss group), and the level of faces (models). I think I just do not understand the difference. Maybe you do not explain it enough in the manuscript, or maybe it is an instance of a think that is "natural to everybody but me."

Furthermore, if the results are reported at the level of the judges and on the level of the judges as discordant, what does this mean about the hypotheses? I would say that the level of the judged (per-face mean - correct me or ignore this notion during the review if I am wrong) is much more important as it is the estimated parameter. Lower/higher correlation on the judges' level is relatively less important.

"The more attractive men appear to be more masculine and more stable; this is certainly a desirable relationship. On the other hand, however, they appear to be less agreeable, less family-oriented, and less cheerful." -> This result is important given that papers from evolutionary psychology recognise that there may exist a dilemma between masculinity on one hand and agreeableness and family orientation on another (e.g., Jones et al. 2019; <https://doi.org/10.1016/j.tics.2018.10.008>). The part "neutralising effect of accounting for height" also presents an interesting outcome.

Some results reports (e.g., page 25/40 in PDF): "When interpreting the data, it is important to remember that the scale does not start at 0, but at 1, so one should mentally subtract one point." could be made more easily accessible.

Throughout the article, you may also consider using more neutral words (instead of, e.g., gigantic, colossal, massive - being used close to each other in a series of sentences). Similarly, statements like "in the world of statistics, these are different universes" caused me to stop reading for a while, simply due to being perplexed. The same is also the case for the beginning of the Discussion and later on ("tragedy"). It would be fun to leave the restricted wordlist "allowed" in a research report (and if there is a call for a revolution, I'm in). However, as a reviewer, I would like to focus on the

information provided, which is hard when all I do is expect another unusual phrasing and terms. Moreover, instances of these unconventional phrases were distributed unevenly; most passages managed to go on well without the need to use such a language.

Moreover, I do not think it is a good decision to start the discussion with notions like the group of raters is non-representative, and the group of stimuli is "a very special group". This rather fits into limitations (that are missing as a separate section). For what reason should I, as a researcher, be interested in a paper with very special stimuli and non-representative raters' sample? Otherwise, based on the obtained results, the discussion reads very well and considers important and interesting aspects of the association between baldness and personal characterisation.

Minor points:

The last point of the Highlights: "The evaluation of attractiveness is very negative, with regard to the sexual component, the verdict is a degradation", is above my skills of reading and understanding English.

Page 9: "This corresponds to an effect size of 0.52" - It is not incorrect to use, but I think the age difference on the scale of years is just enough. Again, rather than helping, this made me stop reading and think, "Why is the effect size so interesting here?"

10/40: "Cronbach's [alpha] and McDoland's [omega] are .99." Choose one measure unless they point to a different favourable characteristic of the dataset with regard to agreement assessment.

Just below: "When analyzing the main experiment, we first look at the level of the judges. With 1618 participants, our sample is very large and hence the test power is very high." Probably, I am just a distracted reader. However, I miss a thorough explanation on what is the level of participants

(sometimes inconsistently referred to as "judges") and level of stimuli ("judged"). Moreover, we can say the power should be very high, but without conducting an analysis, you cannot really say. This paper actually made me start considering why I repeat certain aspects of "why something must be the way it is in research reports". For now, I have decided to insist on these unwritten rules.