

PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (<http://bmjopen.bmj.com/site/about/resources/checklist.pdf>) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Athena SWAN and Gender Diversity: A UK-based Retrospective Cohort Study
AUTHORS	Xiao, Yunyu; Pinkney, Edward; Au, Terry Kit Fong; Yip, Paul

VERSION 1 – REVIEW

REVIEWER	Pavel Ovseiko Senior Research Fellow in Health Policy and Management University of Oxford, UK
REVIEW RETURNED	12-Aug-2019

GENERAL COMMENTS	<p>This is a novel study on an increasingly important topic. The novelty of the study is in using comprehensive HESA data, which hasn't been analysed before in the context of Athena SWAN. In doing so, the study fills a gap in qualitative studies assessing the impact of Athena SWAN. It has a strong potential to inform policy and practice in many countries participating or considering participating in Athena SWAN. I would like to suggest a number of revisions to improve the paper before publication. They largely relate to better explaining the variables and potential limitations of the study. The most important revision is to use data for Athena SWAN awards for the same period as the data for gender diversity.</p> <p>Strengths and Limitations of this Study p.4 This study prospectively investigates – change to retrospectively as per study design This is one of the first studies that examined – use present or past tense consistently Among Leaders in HEIs – capitalise/do not capitalise “Leaders” consistently Athena SWAN levels – perhaps clarify that levels refer to levels of accreditation/award</p> <p>Introduction Despite increasing policies that aims – aim Managed by the Equality Challenge Unit (ECU) since its launch in 2005 – ECU has merged into Advance HE, so perhaps, change to “Athena SWAN was launched in 2005, and is currently managed by Advance HE” The influence and success of the initiative in signing up the majority of HEIs in the UK has inspired similar initiatives internationally, such as in Australia – Ireland joined before Australia, and currently USA and Canada adopted Athena SWAN as well (see https://www.timeshighereducation.com/opinion/athena-swan-taking-flight) It might be useful to add a point about the inclusion of gender in research design and research impact assessment to the background</p>
-------------------------	---

	<p>on the importance of gender in higher education and research: https://health-policy-systems.biomedcentral.com/articles/10.1186/s12961-016-0126-z it might be useful to add to previous research: this recent study analysing Athena SWAN interventions as currently the manuscript doesn't describe Athena SWAN interventions https://doi.org/10.1101/555482 this recent study attributing a positive culture change to Athena SWAN https://doi.org/10.1080/03080188.2019.1603880 this recent viewpoint on unintended consequences https://www.nature.com/articles/d41586-019-01904-1 It might also be useful to mention either in the Introduction or in the Discussion that Athena SWAN is currently under review due to concerns about the award process and administrative burden on institutions: https://www.ecu.ac.uk/equality-charters/athena-swan/review-of-the-athena-swan-charter/</p> <p>Data and Measures Please explain what the different levels of awards require institutions to do as these indicate the extent of intervention and its impact, e.g. if Bronze awards require an assessment of the levels of gender equity and development of an action plan, it is unlikely that the first Bronze awards could reflect changes in the levels of gender equity. Likewise, joining the Athena SWAN charter is not associated with the implementation of gender equity interventions because it only signals an institution's commitment to apply for a Bronze award, i.e. conduct an assessment and develop an action plan. Therefore, analysis can be simplified by treating non-member institutions and non-recipients as a combined control group. Please explain why only institutional level awards were used for this study and how institutional awards differ from departmental awards. In the UK, the role of institutions is rather limited because individual PIs hold budgets and departments make the majority of recruitment and promotion decisions and so the most significant Athena SWAN interventions happen at the departmental level. Also, please explain if institutional awards included awards for research institutes, as these have a slightly different award process. The study analyses data over five years 2012/13-2016/17. Please explain whether institutions had participated/held awards for the same number of years and whether the duration of participation/awards was taken into account. The Discussion suggests that the status of Athena SWAN was based on the most recent records in 2019 and the duration of awards was not taken into account. This is mentioned further in the Limitations. if an institution received its first Bronze award after 2016/17, which is given for the assessment of gender equity and development of an action plan, then it means that during the study period this institution didn't implement any interventions to improve gender equity. If during the study period, one institution implemented two Athena SWAN Silver action plans and another one only one than exposure to interventions in these institutions is different. Unless Athena SWAN awards refer to the same study period, the study cannot claim to have examined the relationship between gender diversity and Athena SWAN awards. I suggest that the authors update their analysis by using the data for Athena SWAN awards corresponding to the study period. The authors can collect the required information from institutional websites because under the Athena SWAN Charter rule institutions commit to publish information about their awards and action plans on their institutional websites. Please explain why only Professors were included in the study and</p>
--	---

	<p>other academic ranks were excluded and that there is an undercount of Professors in F1 category as Professors with institutional or departmental line management responsibilities are included in the Leaders category.</p> <p>If possible, please provide key characteristics of the included institutions and variation between them, e.g. by the number of staff, annual turnover, geography, subject groups (e.g. art/humanities, social sciences, natural sciences, medical sciences, or just STEM vs non-STEMM), and teaching vs research intensity. The latter two characteristics are particularly important because women's advancement and leadership varies by subject and teaching vs research. Research intensity can be measured using HESA data (research income) or through membership in the Russell Group https://russellgroup.ac.uk/about/our-universities/ One possible limitation of the authors' analysis is that it doesn't accounts for differences between HEIs. Comparing results for Russell Group universities and non-RG universities may gain additional important insights. For example, it might be that non-Athena SWAN institutions are non-RG institutions, and so differences in gender diversity between Athena SWAN and non-Athena SWAN institutions might be associated with research vs teaching or STEMM vs non-STEMM focus of the institutions. RG institutions are also presumed to be homogeneous in the levels of excellence and so analysing effects of different levels of Athena SWAN awards in the RG institutions might minimise confounding factors and gain more reliable insights into the impact of Athena SWAN. I would like to encourage authors to consider conducting this additional analysis while appreciating that it is beyond the scope of the current study design.</p> <p>Please briefly explain why QS rankings were selected among other possible rankings and what they measure.</p> <p>Changes of Female Representation over Time As seen in Figure 2, non-Athena institutions had higher percentages of female professors than Athena institutions regardless of whether their Athena levels -- regardless of their Athena SWAN levels of award/accreditation?</p> <p>Discussion Leaders: The study analyses data over five years 2012/13-2016/17. Until 2015, Athena SWAN focused on academic and research staff, but HESA contract levels A-E2 appear to include non-academic/research staff, e.g. Chief Operating Officer/Registrar/Secretary; Director of major function/group of functions e.g. finance, corporate services, HR; Head of Athena SWAN and Gender Diversity; Senior function head. Discuss this as a potential limitation of the study, i.e. changes in the levels of diversity among non-academic/research staff may not be attributable to Athena SWAN.</p> <p>Professors: Professor is the highest academic rank associated with tenured employment, often, once appointed, professorial chairs are held for life and institutions have a limited ability to create new professorial chairs. There is more mobility among lower ranks, and institutions have greater flexibility in appointments and promotion. Arguably, the impact of Athena SWAN could be greater among lower ranks. If the HESA data does not allow examining lower ranks then this needs to be discussed as a limitation/question for future research. Also, Professors with institutional or departmental line management responsibilities are included in the Leaders category and the category F1: Professor (without line management responsibilities) does not include all Professors. Given the divergent</p>
--	---

	<p>results for gender diversity among Leaders and Professors this might be a factor that affected the results for Professors. This, i.e. the undercount of Professors in F1 category, needs to be explained as a limitation of the HESA data and a potential limitation of the overall analysis. If the HESA data allows analysis of lower ranks plus a comprehensive analysis of all Professors, then the authors may want to focus the current paper on the analysis of the Leaders category, and conduct a comprehensive analysis for academic and research staff by different academic ranks in a different paper.</p> <p>Finally, Athena SWAN is a complex intervention which improves conditions for change in the representation in a non-linear way. This paper might be useful for the discussion and interpretation of the results: https://doi.org/10.1101/555482</p> <p>The link below will direct you to the latest evaluation of Athena SWAN which can help you if you incorporate this in your revision: https://www.ecu.ac.uk/publications/an-impact-evaluation-of-the-athena-swan-charter-2019/</p> <p>Overall, I applaud the authors for conducting this study and hope that my comments and suggestions could help the authors to strengthen the paper before publication.</p>
--	---

REVIEWER	Sarah Barnard Loughborough University, UK
REVIEW RETURNED	25-Sep-2019

GENERAL COMMENTS	<p>I read this paper with interest and understand the desire to evaluate the Athena SWAN initiatives as they have an impact internationally on the sector. The front end of the paper reads well, but is very similar to other texts that have been written on Athena SWAN, so some care needs to be taken to avoid plagiarism.</p> <p>I found the use of university rankings somewhat problematic, though I do understand the reasons for going down this route. I would expect there to be a more thorough rationale for this and also acknowledgement of the limitations of this approach. Similarly I found the analysis and discussion to be too simplistic, and lacking linkages to relevant literature. Overall the paper needs a lot of work to get to publishable standard.</p> <p>Detailed comments:</p> <ul style="list-style-type: none"> - Abstract - you need to define what you mean by university performance - The research questions (and hypotheses) should be more clearly outlined. - Some consideration of the difference between departmental and institutional awards, and the complexity of comparing institutions on the basis of one (institution level), but not the other (department level). - You do not consider fully the relationship between Athena SWAN and gender equality progress. The main thrust is that you are trying to understand the impact of Athena SWAN on gender equality through the data analysis. However, it should be considered how far institutions that are dominated by STEMM subjects (particularly medicine) are under pressure to take action and the extent to which those that are making progress in gender equality have impetus to apply for the higher awards. The awards are meant to recognize both qualitative and quantitative progress and therefore the
-------------------------	---

	<p>relationship between representation data and awards is not as simple as suggested in the paper.</p> <ul style="list-style-type: none"> - Methods section - the female to male ratios are stated as being rounded up. I would think that you either round up before working out ratios or don't do it at all? Please check the correct way of applying data protection laws. - Methods section - I am not sure it is necessary to list all the coding/categories of staff in the paper. - Page 12, 3rd para. First sentence is unclear.
--	--

REVIEWER	Shahadat Uddin The University of Sydney, Australia
REVIEW RETURNED	01-Oct-2019

GENERAL COMMENTS	<p>This review is entirely based on the statistical methods and concepts used in this manuscript.</p> <ol style="list-style-type: none"> 1. Table 1: the way authors calculate the 'rate of change' is ambiguous. It could be the case when there is no change in years 2013 and 2017 but big changes in the middle years (2014-2016). I would suggest the authors to follow the keyword growth concept on this article - https://doi.org/10.1016/j.joi.2016.10.004. The authors of this article developed a similar measure in this respect but applied it to a different context. 2. From a quick understanding of M and SD values, the statistical tests (t-test) of tables 2-3 look okay. 3. The only mistake is that for 'professors' the value for n should be 102. At present, it says (in the caption of table 2) this value is 103. 4. I don't see the necessity of table 3 since this table shows the result of only one t-test. The same result can be described using one sentence.
-------------------------	---

VERSION 1 – AUTHOR RESPONSE

Reviewer 1

Comments	Responses
<p>This is a novel study on an increasingly important topic. The novelty of the study is in using comprehensive HESA data, which hasn't been analysed before in the context of Athena SWAN. In doing so, the study fills a gap in qualitative studies assessing the impact of Athena SWAN. It has a strong potential to inform policy and practice in many countries participating or considering participating in Athena SWAN. I would like to suggest a number of revisions to improve the paper before publication. They largely relate to better explaining the variables and potential limitations of the study. The most important revision is to use data for Athena SWAN awards for the same period as the data for gender diversity.</p>	<p>Thank you very much for your positive feedback and constructive suggestions. We have revised the manuscript by comprehensively addressing each of the comments. The revised paper is a considerable improvement after incorporating the recommended changes.</p>
Strengths and Limitations of this Study	

<p>p.4 This study prospectively investigates – change to retrospectively as per study design</p> <p>This is one of the first studies that examined – use present or past tense consistently</p> <p>Among Leaders in HEIs – capitalise/do not capitalise “Leaders” consistently</p> <p>Athena SWAN levels – perhaps clarify that levels refer to levels of accreditation/award</p>	<p>Thank you for the suggestions, we have revised this section according to address each point (See page 4).</p>
<p>The introduction:</p>	
<p>Despite increasing policies that aims – aim Managed by the Equality Challenge Unit (ECU) since its launch in 2005 – ECU has merged into Advance HE, so perhaps, change to “Athena SWAN was launched in 2005, and is currently managed by Advance HE”</p> <p>The influence and success of the initiative in signing up the majority of HEIs in the UK has inspired similar initiatives internationally, such as in Australia – Ireland joined before Australia, and currently USA and Canada adopted Athena SWAN as well (see https://urldefense.proofpoint.com/v2/url?u=https-3A__www.timeshighereducation.com_opinion_athen-2Dswan-2Dtaking-2Dflight&d=DwIFaQ&c=slrrB7dE8n7gBJbeO0g-IQ&r=RrPbLBpCLkw-FTSez5y_UQ&m=vKURJKVRp8D3Ht03A0J3zZ0C-3QpiivBKSwLSW4r7mE&s=V586vWayOj7AHcSoJUNk2BmL41CseFMsfbVeNN01YQM&e=))</p>	<p>Thank you for the suggestions, we have revised this section accordingly and provided background of the influence and early success of this initiative (See page 5).</p>
<p>It might be useful to add a point about the inclusion of gender in research design and research impact assessment to the background on the importance of gender in higher education and research: https://urldefense.proofpoint.com/v2/url?u=https-3A__health-2Dpolicy-2Dsystems.biomedcentral.com_articles_10.1186_s12961-2D016-2D0126-2Dz&d=DwIFaQ&c=slrrB7dE8n7gBJbeO0g-IQ&r=RrPbLBpCLkw-FTSez5y_UQ&m=vKURJKVRp8D3Ht03A0J3zZ0C-3QpiivBKSwLSW4r7mE&s=LpbzOF0hXMCizS8J-YtJAUwwBMzxO_t2KAJxdy8tGls&e=</p> <p>it might be useful to add to previous research:</p> <p>this recent study analysing Athena SWAN interventions as currently the manuscript doesn’t describe Athena SWAN interventions https://urldefense.proofpoint.com/v2/url?u=https-3A__doi.org_10.1101_555482&d=DwIFaQ&c=slrrB7dE8n7gBJbeO0g-IQ&r=RrPbLBpCLkw-FTSez5y_UQ&m=vKURJKVRp8D3Ht03A0J3zZ0C-3QpiivBKSwLSW4r7mE&s=Ke4DARVjnuhSFS-LNMOM71FB7meOb6YnnRKcXf9VdXQ&e=</p> <p>this recent study attributing a positive culture change to Athena SWAN https://urldefense.proofpoint.com/v2/url?u=https-3A__doi.org_10.1080_03080188.2019.1603880&d=DwIFaQ&c=slrrB7dE8n7gBJbeO0g-IQ&r=RrPbLBpCLkw-FTSez5y_UQ&m=vKURJKVRp8D3Ht03A0J3zZ0C-3QpiivBKSwLSW4r7mE&s=-YWyHkwdCJX7vzjwnfpdssXLmOehleCIKwbik6pytJ0&e=</p> <p>this recent viewpoint on unintended consequences https://urldefense.proofpoint.com/v2/url?u=https-3A__www.nature.com_articles_d41586-2D019-2D01904-2D1&d=DwIFaQ&c=slrrB7dE8n7gBJbeO0g-IQ&r=RrPbLBpCLkw-</p>	<p>Thank you for your suggestion and detailed reference, we have added the points about the influences of including gender in research and relevant discussion on Athena SWAN interventions throughout the Introduction (see pages 5-10) and addressed the concerns about the award process in the discussion (see pages 18-21).</p>

<p>FTSez5y_UQ&m=vKURJKVRp8D3Ht03A0J3zZ0C-3QpiivBKSwLSW4r7mE&s=A8yVMwYI4iMvLFtM63hdl7Uj0Dmn_VPpd9coabzPNtk&e=</p> <p>It might also be useful to mention either in the Introduction or in the Discussion that Athena SWAN is currently under review due to concerns about the award process and administrative burden on institutions: https://urldefense.proofpoint.com/v2/url?u=https-3A__www.ecu.ac.uk_equality-2Dcharters_athena-2Dswan_review-2Dof-2Dthe-2Dathena-2Dswan-2Dcharter_&d=DwlFaQ&c=slrrB7dE8n7gBJbeO0g-IQ&r=RrPbLBpCLkw-FTSez5y_UQ&m=vKURJKVRp8D3Ht03A0J3zZ0C-3QpiivBKSwLSW4r7mE&s=8_yvc3eD0w2s8bE1IXtrKDBDOeZon-vfJTzQutVV8DE&e=</p>	
<p>Data and Measures:</p>	
<p>Please explain what the different levels of awards require institutions to do as these indicate the extent of intervention and its impact, e.g. if Bronze awards require an assessment of the levels of gender equity and development of an action plan, it is unlikely that the first Bronze awards could reflect changes in the levels of gender equity. Likewise, joining the Athena SWAN charter is not associated with the implementation of gender equity interventions because it only signals an institution's commitment to apply for a Bronze award, i.e. conduct an assessment and develop an action plan. Therefore, analysis can be simplified by treating non-member institutions and non-recipients as a combined control group.</p>	<p>Thank you for your insightful comments.</p> <p>We now explain in details regarding the requirements of institutions to receive Bronze and Silver awards (see page 12).</p> <p>In our analysis, we separated the non-member institutions and non-recipients.</p>
<p>Please explain why only institutional level awards were used for this study and how institutional awards differ from departmental awards. In the UK, the role of institutions is rather limited because individual PIs hold budgets and departments make the majority of recruitment and promotion decisions and so the most significant Athena SWAN interventions happen at the departmental level. Also, please explain if institutional awards included awards for research institutes, as these have a slightly different award process.</p>	<p>Thanks for the great advice.</p> <p>We did not use the departmental awards since this research purpose of this study is to examining gender diversity at the institutional levels.</p> <p>We further explain the differences between institutional and departmental awards (see page 12).</p>
<p>The study analyses data over five years 2012/13-2016/17. Please explain whether institutions had participated/held awards for the same number of years and whether the duration of participation/awards was taken into account. The Discussion suggests that the status of Athena SWAN was based on the most recent records in 2019 and the duration of awards was not taken into account. This is mentioned further in the Limitations. if an institution received its first Bronze award after 2016/17, which is given for the assessment of gender equity and development of an action plan, then it means that during the study period this institution didn't implement any interventions to improve gender equity. If during the study period, one institution implemented two Athena SWAN Silver action plans and another one only one than exposure to interventions in these institutions is different. Unless Athena SWAN awards refer to the same study period, the study cannot claim to have examined the relationship between gender</p>	<p>Thank you for your suggestions.</p> <p>We have updated our analysis by using the data for Athena SWAN awards corresponding to the study period.</p> <p>We collected the required information from institutional websites as well as the</p>

<p>diversity and Athena SWAN awards. I suggest that the authors update their analysis by using the data for Athena SWAN awards corresponding to the study period. The authors can collect the required information from institutional websites because under the Athena SWAN Charter rule institutions commit to publish information about their awards and action plans on their institutional websites.</p>	<p>annual reports from ECU.</p> <p>We also create a variable to account for the duration of awards.</p> <p>The Methods (See pages 11-13) and Results See pages 15-18) sections have been updated to reflect these changes.</p>
<p>Please explain why only Professors were included in the study and other academic ranks were excluded and that there is an undercount of Professors in F1 category as Professors with institutional or departmental line management responsibilities are included in the Leaders category.</p>	<p>Thanks very much for your inquiries. This study focuses on the senior levels of administrative (Leaders) and academic (Professors). Therefore, faculty members with lower academic ranking other than Professor were excluded.</p> <p>We defined Leaders category as senior members with management responsibilities. Hence, we included Professors institutional or departmental line management responsibilities in our Leaders category. These are explained in the Method section (See pages 10-11).</p> <p>We recognize the meaningfulness to assess effects of Athena on the lower-rank professor category. Hence, we added it as a future direction in our Discussion (See pages 21-22).</p>
<p>If possible, please provide key characteristics of the included institutions and variation between them, e.g. by the number of staff, annual turnover, geography, subject groups (e.g. art/humanities, social sciences, natural sciences, medical sciences, or just STEM vs non-STEM), and teaching vs research intensity. The latter two characteristics are particularly important because women's advancement and leadership varies by subject and teaching vs research. Research intensity can be measured using HESA data (research income) or through membership in the Russell Group https://urldefense.proofpoint.com/v2/url?u=https-3A__russellgroup.ac.uk_about_our-</p>	<p>Thank you for your insights.</p> <p>We agree that these are interesting and meaningful analyses for further investigation. However, since this study focuses on Athena SWAN</p>

<p>2Duniversities_&d=DwIFaQ&c=slrrB7dE8n7gBJbeO0g-IQ&r=RrPbLBpCLkw-FTSez5y_UQ&m=vKURJKVRp8D3Ht03A0J3zZ0C-3QpiivBKSwLSW4r7mE&s=MoWtDIGTW5sKGSVVSzIZO5UAfTvXn9jHt6CMhsRJw04&e= One possible limitation of the authors' analysis is that it doesn't accounts for differences between HEIs. Comparing results for Russell Group universities and non-RG universities may gain additional important insights. For example, it might be that non-Athena SWAN institutions are non-RG institutions, and so differences in gender diversity between Athena SWAN and non-Athena SWAN institutions might be associated with research vs teaching or STEMM vs non-STEMM focus of the institutions. RG institutions are also presumed to be homogeneous in the levels of excellence and so analysing effects of different levels of Athena SWAN awards in the RG institutions might minimise confounding factors and gain more reliable insights into the impact of Athena SWAN. I would like to encourage authors to consider conducting this additional analysis while appreciating that it is beyond the scope of the current study design.</p>	<p>accreditation/awards and female representation, as well as university performance, the suggested analyses are beyond the scope of the current study design. We feel encouraged to conduct a follow-up study based on this advice. We have further addressed this point in our Discussion (see pages 21-22).</p>
<p>Please briefly explain why QS rankings were selected among other possible rankings and what they measure.</p>	<p>Thanks very much for the comments. We add more information about different measures of world university ranking system and explain the reasons to choose QS rankings in the Data and Measures section (See pages 13).</p>
<p>Changes of Female Representation over Time</p>	
<p>As seen in Figure 2, non-Athena institutions had higher percentages of female professors than Athena institutions regardless of whether their Athena levels -- regardless of their Athena SWAN levels of award/accreditation?</p>	<p>Thank you very much for your suggestions. We have revised this sentence accordingly (see page 16-17).</p>
<p>Discussion</p>	
<p>Leaders: The study analyses data over five years 2012/13-2016/17. Until 2015, Athena SWAN focused on academic and research staff, but HESA contract levels A-E2 appear to include non-academic/research staff, e.g. Chief Operating Officer/Registrar/Secretary; Director of major function/group of functions e.g. finance, corporate services, HR; Head of Athena SWAN and Gender Diversity; Senior function head. Discuss this as a potential limitation of the study, i.e. changes in the levels of diversity among non-academic/research staff may not be attributable to Athena SWAN.</p>	<p>Thank you for your insights. We note this point as the fourth limitation in the revised manuscript (see pages 21-22).</p>

<p>Professors: Professor is the highest academic rank associated with tenured employment, often, once appointed, professorial chairs are held for life and institutions have a limited ability to create new professorial chairs. There is more mobility among lower ranks, and institutions have greater flexibility in appointments and promotion. Arguably, the impact of Athena SWAN could be greater among lower ranks. If the HESA data does not allow examining lower ranks then this needs to be discussed as a limitation/question for future research. Also, Professors with institutional or departmental line management responsibilities are included in the Leaders category and the category F1: Professor (without line management responsibilities) does not include all Professors. Given the divergent results for gender diversity among Leaders and Professors this might be a factor that affected the results for Professors. This, i.e. the undercount of Professors in F1 category, needs to be explained as a limitation of the HESA data and a potential limitation of the overall analysis. If the HESA data allows analysis of lower ranks plus a comprehensive analysis of all Professors, then the authors may want to focus the current paper on the analysis of the Leaders category, and conduct a comprehensive analysis for academic and research staff by different academic ranks in a different paper.</p>	<p>Thank you for your thoughtful comment.</p> <p>It is possible that the impact of Athena SWAN could be concentrated on or greater among lower-rank professors. However, the study objective is to examine the impact of Athena SWAN in senior-level administrative (managerial) and academic positions. Hence, we did not include the analysis for lower-rank professors and would prefer to conduct that analysis in a different paper.</p> <p>We include this point as a limitation and opportunity for future study (in the “Fifth” limitation, see page 22).</p> <p>Thank you for your suggestion. However, the HESA data only provides either 1) combined A-E2 or 2) F1, which does not allow a comprehensive analysis of all Professors. We feel encouraged to collect further data and conduct a different paper.</p>
<p>Finally, Athena SWAN is a complex intervention which improves conditions for change in the representation in a non-linear way. This paper might be useful for the discussion and interpretation of the results: https://urldefense.proofpoint.com/v2/url?u=https-3A__doi.org_10.1101_555482&d=DwIFaQ&c=slrrB7dE8n7gBJbeO0g-IQ&r=RrPbLBpCLkw-FTSez5y_UQ&m=vKURJKVRp8D3Ht03A0J3zZ0C-3QpiivBKSwLSW4r7mE&s=Ke4DARVjnuhSFS-LNMOM71FB7meOb6YnnRKcXf9VdXQ&e=</p>	<p>Thank you for the helpful advice and wonderful resource. We add this resource in the discussion (see page 21).</p>
<p>The link below will direct you to the latest evaluation of Athena SWAN which can help you if you incorporate this in your revision: https://urldefense.proofpoint.com/v2/url?u=https-3A__www.ecu.ac.uk_publications_an-2Dimpact-2Devaluation-2Dof-2Dthe-2Dathena-2Dswan-2Dcharter-2D2019_&d=DwIFaQ&c=slrrB7dE8n7gBJbeO0g-IQ&r=RrPbLBpCLkw-FTSez5y_UQ&m=vKURJKVRp8D3Ht03A0J3zZ0C-3QpiivBKSwLSW4r7mE&s=jgdChW4hQKqH8bglixRwYXzoYyDoZ21dyPcW54cHgoQ&e=</p>	<p>Thank you very much for the great resource. We have include it in our revision (see page 21).</p>

Conclusion	
Overall, I applaud the authors for conducting this study and hope that my comments and suggestions could help the authors to strengthen the paper before publication.	Thank you very much for your interests in our paper and constructive suggestions. We believe the paper has been improved significantly for publication.

Reviewer 2

Comments	Responses
I read this paper with interest and understand the desire to evaluate the Athena SWAN initiatives as they have an impact internationally on the sector. The front end of the paper reads well, but is very similar to other texts that have been written on Athena SWAN, so some care needs to be taken to avoid plagiarism.	Thank you for your positive feedback of our paper. We have revised the manuscript by comprehensively addressing each of the comments. We specifically pay attention to avoid similar texts in the previous Athena SWAN report. The revised paper is a considerable improvement after incorporating the suggested changes
I found the use of university rankings somewhat problematic, though I do understand the reasons for going down this route. I would expect there to be a more thorough rationale for this and also acknowledgement of the limitations of this approach. Similarly I found the analysis and discussion to be too simplistic, and lacking linkages to relevant literature. Overall the paper needs a lot of work to get to publishable standard.	Thanks for the comment. We add a more thorough rationale for using university ranking (see pages 13), and acknowledge possible limitations of this approach (see page 22).
Abstract	
You need to define what you mean by university performance	Thank you for the advice. We now state the meaning of university performance in the abstract. -- as measured by overall scores in the global ranking systems (see page 2).
Introduction	
- The research questions (and hypotheses) should be more clearly outlined.	These are excellent points to clarify. We revise the research questions and hypotheses to be clearer (see pages 9-10).

<p>- Some consideration of the difference between departmental and institutional awards, and the complexity of comparing institutions on the basis of one (institution level), but not the other (department level).</p>	<p>Thanks for the comment. As an exploratory study, we focus on institutional-level awards to be more comparable across institutions. We further clarify this point in pages 11-12.</p>
<p>- You do not consider fully the relationship between Athena SWAN and gender equality progress. The main thrust is that you are trying to understand the impact of Athena SWAN on gender equality through the data analysis. However, it should be considered how far institutions that are dominated by STEMM subjects (particularly medicine) are under pressure to take action and the extent to which those that are making progress in gender equality have impetus to apply for the higher awards. The awards are meant to recognize both qualitative and quantitative progress and therefore the relationship between representation data and awards is not as simple as suggested in the paper.</p>	<p>Thanks very much for the suggestions. We acknowledge that both qualitative and quantitative progress shall be evaluated. While previous studies are mostly focusing on qualitative examination, our study contributes to the literature by adding a quantitative perspective. We add the need for considering both approaches in the Discussion (see page 22).</p>
<p>Method</p>	
<p>the female to male ratios are stated as being rounded up. I would think that you either round up before working out ratios or don't do it at all? Please check the correct way of applying data protection laws.</p>	<p>Thanks very much for your inquiry. The ratios were rounded up in the original data provided by HESA.</p>
<p>I am not sure it is necessary to list all the coding/categories of staff in the paper.</p>	<p>Thank you for your insights. We decide to keep the list as it facilitates discussion about the complex influence of Athena SWAN in different levels of staffs in HEIs (see page 10-11, 21-22).</p>
<p>- Page 12, 3rd para. First sentence is unclear.</p>	<p>We revise the sentence to be clearer. Thank you!</p>

Reviewer 3

Comments	Responses
----------	-----------

<p>This review is entirely based on the statistical methods and concepts used in this manuscript.</p>	<p>Thank you for your insightful comments on our statistical methods and concepts.</p>
<p>1. Table 1: the way authors calculate the 'rate of change' is ambiguous. It could be the case when there is no change in years 2013 and 2017 but big changes in the middle years (2014-2016). I would suggest the authors to follow the keyword growth concept on this article - https://urldefense.proofpoint.com/v2/url?u=https-3A__doi.org_10.1016_j.joi.2016.10.004&d=DwIFaQ&c=slrrB7dE8n7gBJbeO0g-IQ&r=RrPbLBpCLkw-FTSez5y_UQ&m=vKURJKVRp8D3Ht03A0J3zZ0C-3QpiiivBKSwLSW4r7mE&s=N5bW5syRGxDBQHY6lwUGgYnkCDhj63QEG_qrOsKI4sl&e= . The authors of this article developed a similar measure in this respect but applied it to a different context.</p>	<p>Thanks for the excellent feedback and resources. We follow the 'keyword growth' concept and revise our calculation of 'rate of change' (now rephrased as 'rate growth'). We describe the method to calculate our rate growth in detail in the Method section (see pages 13-14). The updated figures are shown in Table 1 and discussed throughout the relevant Result sections (see pages 15-16).</p>
<p>2. From a quick understanding of M and SD values, the statistical tests (t-test) of tables 2-3 look okay.</p>	<p>Thank you for helping us examine these values.</p>
<p>3. The only mistake is that for 'professors' the value for n should be 102. At present, it says (in the caption of table 2) this value is 103.</p>	<p>Thank you for the useful feedback. We correct it in Table 2.</p>
<p>4. I don't see the necessity of table 3 since this table shows the result of only one t-test. The same result can be described using one sentence.</p>	<p>Thank you for your comments. We decided to delete this table accordingly and replaced it with a more comprehensive table capturing the relationship between Athena award and QS ranking overall scores.</p>

VERSION 2 – REVIEW

<p>REVIEWER</p>	<p>Pavel Ovseiko Senior Research Fellow in Health Policy and Management University of Oxford, UK</p>
<p>REVIEW RETURNED</p>	<p>23-Dec-2019</p>
<p>GENERAL COMMENTS</p>	<p>The authors have critically considered and comprehensively addressed my comments and suggestions. I would like to recommend this manuscript for publication and encourage the authors to continue their research on this important topic.</p>