**Response to Reviewers’ Comments**

**Date**: January 8, 2023

**Manuscript Number**: SCIM-D-22-00945R1

**Title of Article**: Temporal trends in academic performance and career duration of principal investigators in ecology and evolutionary biology in Taiwan

**Corresponding Author**: Syuan-Jyun Sun

**Email**: sjs243@ntu.edu.tw

---------------------------------------------------------------------------------------------------------------------

Dear Dr. Lin Zhang,

Thank you for inviting us to submit a revised version of the manuscript. We greatly appreciate the valuable comments and feedback from the reviewers. We have incorporated many of the suggestions and the revision has substantially improved the manuscript. In particular, we have made the following major changes:

* Revised/added two paragraphs in the introduction section for a review of past studies on this topic to provide more background information regarding the current state of the research.
* Added a paragraph in the methods section detailing the searching process and data handling in the *Publish or Perish* software.
* Added a paragraph in the methods section discussing the advantages and disadvantages of Google Scholar Profile.
* Added a paragraph in the methods section discussing the advantages and disadvantages of h-index.
* Update Line numbers
* Re-analyzed the data and updated the statistical analyses, results, and discussion.
* Added a concluding paragraph in the discussion section providing practical advice and suggestions for people who hope to pursue an academic career and who are currently progressing through their career stages.
* Modified Table 1 to provide more details on the models results and fit measures (model specifications, sample sizes, regression coefficients, standard errors, 95% confidence limits).

Please also see the following section for our detailed point-by-point responses. All line numbers pertaining to the changes refer to the revised manuscript.

Sincerely,

Syuan-Jyun Sun

---------------------------------------------------------------------------------------------------------------------

**Reviewer 1's Comments to the Author(s):**

This article collects various types of open data and investigates the academic job market in Taiwan, which is an important topic deserving further future studies. However, there are some questions or suggestions as follows:

*(1) About data collection*

**Comment 1** > In line 123 to 126, this research includes 145 PIs who had an updated CV. In other words, researchers who hadn't updated CV were excluded and those PIs' academic careers whether represent specific patterns or not. Hence, what's the representativeness of these 145 PIs in this study?

**Response** >

In most cases, we were able to obtain the PhD education as well as year of recruitment/promotion of PIs on their personal or department/institute websites (these are the most basic information of researchers’ profile). For those PIs with missing information on the websites, we would further search their profile online and record the necessary information to complete our data. The final data include PIs from seven major universities and one research institute; therefore, we feel that our data should be fairly representative of the PIs in ecology and evolutionary biology in Taiwan.

**Comment 2** > Besides the institutional/departmental websites, and ORCID, other open data or database could be further consideration, such as Web of Science, Scopus, or Academic Research Service Portal Researcher Query of National Science and Technology Council.

**Response** > Thanks for the suggestion. Yes, these websites are great sources for PIs’ profile. Indeed, as mentioned in the response to Comment 1, we were able to obtain the necessary information for most PIs through their personal or department/institute websites and ORCID sites.

1. *About literatures*

**Comment 3** > This study includes those variables such as year of recruitment, gender, PhD university origin, PhD university ranking, year of promotion and so on (shown as Table 1). What's the theoretical basis of relationship between those variables? If this article supplied the section of literature review, readers would more understand the existing related researches of this topic even the theoretical basis.

**Response** > Thanks for the suggestion. We agree that it would be helpful to provide some background information so that the readers can get a better idea of the research context. We have now included the following two paragraphs (Line 84-92; Line 94-108) in the introduction section:

*“Previous studies have focused on how various bibliometric indicators predict researchers’ future academic excellence and scientific contributions. The number of publications, top journal publications, publication rates, the number of distinct journals, and the number of citations are all important determinants of academic performance (Acuna et al. 2012; Danell 2011; Lindahl 2018). Academic performance is critical for researchers’ future success as publication requirements for recruitment as a new faculty member and promotion to full professor have surged in recent years, yet empirical quantification of how performance affects the duration before recruitment and promotion over time remains unexplored.”*

*”In addition to research performance, the prestige of doctoral-granting institutes are critical indicators for academic employment as well (van Dijk et al. 2014), especially in East Asian countries (Shin and Kehm 2013). With the initiative to build world-class universities, many East Asian universities preferentially recruit returnees who obtained PhD degrees from top-ranked universities in Western countries. Hence, competition for limited tenure-track positions is exacerbated when foreign PhDs are favored, leaving domestically-trained PhDs deprived of career development opportunities (Chen 2021). However, whether and to what extent publication performance and career duration differ between researchers with domestic and foreign degrees, and whether their pre- and post-employment performance changes, remain largely unexplored. Moreover, studies have shown that the researcher’s gender determines the probability of becoming a principal investigator (PI) (van Dijk et al. 2014) and receiving grants (Witteman et al. 2019), yet little is known about how gender affects the publication performance and career duration for recruitment and promotion.”*

**Comment 4** > According to the results and discussions, what's the concrete suggestions to higher education policy, recruitment of university's teaching and research staffs, or PhD students who aim to academic careers?

**Response** > Thanks for this critical comment. A major goal of our study is to provide practical advice for people in the academic community. We have now added the following concluding paragraph in the discussion section for this (Line 374-386):

*“Taken together, our study confirms that succeeding in academia has become more challenging, with performance requirements and career duration both increasing over years. Based on our findings, we provide several suggestions for people who hope to pursue an academic career and who are progressing through their career stages: (1) For PhD students and early-career researchers, focusing on research performance may facilitate future academic success, but other aspects of academics (e.g., scientific communication and networking) are important as well. (2) For researchers who have landed a position, fulfilling institute’s requirements while maintaining academic outputs may accelerate the promotion process. (3) For researchers with domestic degrees, seeking international collaboration to expand research network may help enhance productivity. Finally, regardless of career stage, boosting performance is the ultimate key to academic success in the face of increasingly competitive academic job markets.”*

**Reviewer 2's Comments to the Author(s):**

**Comment 1** > The authors examine "how academic performance as well as duration before recruitment as a new principal inverstigator (PI) and promotion to full professor changed over time, and how PhD university origin, PhD university ranking, and gender affected the career success". The manuscript has potential to make a contribution to the literature. However, the manuscript has some problems which makes me recommend major revision. Hopefully my questions and comments can help the authors to improve the manuscript.

**Response** > Thanks for the positive attitude towards this study. We have revised our manuscript based on the following comments and incorporated many of the suggestions provided.

*Title*

**Comment 2** > The title does not really convey what the study is about.

**Response** > We have now changed our title after the revision: *”Temporal trends in academic performance and career duration of principal investigators in ecology and evolutionary biology in Taiwan”*.

*Abstract*

**Comment 3** > The first sentence in the abstract "Academic job markets have become increasingly challenging worldwide, yet it remains poorly characterized how competitively-successful candidates should be and what the underlying determinants of their success are" seem unsubstantiated. There are numerous studies that have examined determinants of academic success (see. e.g., Hirsch, 2007; Danell, 2011; Acuna et al., 2012; Havemann and Larsen, 2015; Bornmann and Williams, 2017a; Lindahl, 2018). My recommendation is that the authors include a more extensive literature review on previous research in the field and provide a more accurate and nuanced summary of the state of this research.

**Response** > Thanks for the suggestion. We have now revised the first part of the abstract and revised/added two paragraphs in the introduction section to provide more background information of the topic so that the readers can get a better idea of the research context (also see our response to Comment 3 from Reviewer 1).

* Revised abstract (Line 27-31):

*“Academic job markets have become increasingly challenging worldwide, with rising performance requirements for recruitment as a new faculty member and promotion to full professor in recent years. However, it remains underexplored how research performance and other determinants of academic success, including PhD university origin, prestige, and gender, affect recruitment and promotion over time.”*

* Revised/added paragraphs in the introduction section (Line 84-92; Line 94-108):

*“Previous studies have focused on how various bibliometric indicators predict researchers’ future academic excellence and scientific contributions. The number of publications, top journal publications, publication rates, the number of distinct journals, and the number of citations are all important determinants of academic performance (Acuna et al. 2012; Danell 2011; Lindahl 2018). Academic performance is critical for researchers’ future success as publication requirements for recruitment as a new faculty member and promotion to full professor have surged in recent years, yet empirical quantification of how performance affects the duration before recruitment and promotion over time remains unexplored.”*

*”In addition to research performance, the prestige of doctoral-granting institutes are critical indicators for academic employment as well (van Dijk et al. 2014), especially in East Asian countries (Shin and Kehm 2013). With the initiative to build world-class universities, many East Asian universities preferentially recruit returnees who obtained PhD degrees from top-ranked universities in Western countries. Hence, competition for limited tenure-track positions is exacerbated when foreign PhDs are favored, leaving domestically-trained PhDs deprived of career development opportunities (Chen 2021). However, whether and to what extent publication performance and career duration differ between researchers with domestic and foreign degrees, and whether their pre- and post-employment performance changes, remain largely unexplored. Moreover, studies have shown that the researcher’s gender determines the probability of becoming a principal investigator (PI) (van Dijk et al. 2014) and receiving grants (Witteman et al. 2019), yet little is known about how gender affects the publication performance and career duration for recruitment and promotion.”*

*Materials and Methods*

Measurement of academic performance

**Comment 4** > The data collection with the Publish or Perish software for the h-index need to be described and presented much more and in greater detail. There is not enough detail to be able to review the data collection or the data for calculating the h-index. As a reader I'm not sure how the authors collected the publications for the authors. Did they conduct searches through Publish or Perish at the publication level or the author level? How was the search queries formulated, i.e., did the authors conduct searches on the basis of publication titles, persistent identifiers, etc? My recommendation is (1) that the authors provide the search queries in the manuscript or as an appendix and (2) that they provide much more detail about the data collection procedure and what they have done including how they handle the CV data etcetera.

**Response** > We have added a paragraph in the methods section detailing our searching process and data handling in the *Publish or Perish* software (Line 148-163):

*“We collected citation data on PIs via the Publish or Perish software, which uses Google Scholar Profile queries to obtain citation information of researchers’ publications and converts it into several citation metrics (e.g., total number of citations, citations per year, and h-index). The data collection was conducted at the author level by entering each PI’s full name or the abbreviated version in scientific publications to the search field. The range of years was set based on the year of recruitment and promotion for each PI (five-year interval before and after the year of recruitment/promotion; see the following section Measurement of academic performance for more details). After the search was completed, we checked individually each publication item in the results pane and included only peer-reviewed papers and book chapters regardless of authorship (PhD theses and conference presentations were excluded). Separate result items that indeed referred to a single article or a book were merged into one entry. We also cross-referenced the result items with the updated curriculum vitae online to ensure the accuracy of search results. The final citation metrics were then exported for further statistical analyses.”*

**Comment 5** > I cannot see how many documents that are included in the final dataset? This should be included in the manuscript.

**Response** > As mentioned in the methods section, our final dataset includes 145 PIs from seven major universities and one research institute in Taiwan. The data compilation process consisted of two phases:

Phase 1: We visited the department/institute websites and/or ORCID sites to obtain information on the PhD education and year of recruitment/promotion of PIs in ecology and evolutionary biology. 145 PIs were identified and recorded in the phase.

Phase 2: For each of those 145 PIs, we entered the name in the *Publish or Perish* software to retrieve the publication information and related citation metrics, exported the results, and combined them with the data obtained in Phase 1.

There are no actual “documents” in the final dataset, but rather it incorporates the PIs’ information obtained from the websites as well as the *Publish or Perish* software.

**Comment 6** > What do the authors mean with "regardless of authorship for" in the sentence on page 7 row 1-3?

**Response** > By "regardless of authorship for", we mean that we included any publications of the PI for the calculation of h-index, regardless of whether the PI was the first author, co-author, or corresponding author. We have changed the original expression to "regardless of authorship” (Line 158).

**Comment 7** > The authors use the h-index to measure research performance. The h-index is not a normalized indicator of research performance, i.e., it do not adjust for, e.g., research area, publication year, and publication type, and do not live up to best practice in scientometric research. See e.g., Waltman (2016) for a review of citation indicators. To use non-normalized bibliometric indicators as measures of research performance can lead to severe biases in the analyses. The h-index has been heavily criticized in the scientometric literature (Bornmann, & Daniel, 2007; Bornmann, & Daniel, 2009) and it is not recommended to use to measure research performance at the individual level (Waltman, & Van Eck, 2012). My recommendation is that the authors change their dependent variable to a normalized bibliometric indicator that is in accordance with best practice in scientometric research or provide good arguments for why the use of h-index should be used in this case. Another potential solution is to use a variation of the h-index that adjust for the problems with the h-index and fit the context of the authors study (see e.g., Alonso et al., 2009, for a review of h-index and its variant).

**Response** > Thanks for pointing out this issue. We agree that there are several drawbacks of h-index. However, we still decided to use it in our re-analyses for the following reasons:

* **Advantages of h-index:**

1. h-index is robust to highly- and lowly-cited publications and thus is suitable for evaluating the overall impact of a researcher’s outputs (Bornmann and Daniel 2007).
2. The original h-index is more straightforward and readily available on various academic search engines compared to the alternative normalized versions.

* **Concerns about h-index (Waltman 2016):**

1. h-index varies among research areas: Our study focuses on PIs within the field of ecology and evolutionary biology and thus their h-indexes should be fairly comparable.
2. h-index depends on the publication year:
3. h-index varies among publication type: We included both journal articles and book (chapters) in the calculation of h-index rather than calculating h-indexes separately for different publication types.

We have included a discussion of the use of h-index in our study in the methods section (Line 183-188):

*“Furthermore, h-index is robust to a few highly-cited or a set of lowly-cited publications, rending it suitable for evaluating the overall impact of a researcher’s outputs (Bornmann and Daniel 2007). Although h-index can vary considerably among different fields of study (Alonso et al. 2009), we focused on PIs within the field of ecology and evolutionary biology and thus their h-indexes should be fairly comparable.”*

**References:**

Bornmann, L. & Daniel, H. D. (2007). What do we know about the h index?. *Journal of the American Society for Information Science and technology*, 58, 1381-1385.

Waltman, L. (2016). A review of the literature on citation impact indicators. *Journal of informetrics*, 10(2), 365-391.

**Comment 8** > The authors need to discuss the pros and cons of using google scholar. Why use Google Scholar instead of a citation database, e.g., Scopus or the citation indices accessible through Web or Science? Harzing state that Web of Science and Scopus have higher accuracy so why not use them (see [https://harzing.com/resources/publish-or-perish/manual/using/query-results/accuracy)?](https://harzing.com/resources/publish-or-perish/manual/using/query-results/accuracy)?" \t "_blank) How might the use of Google Scholar affect the results? The authors should provide a discussion in the manuscript where the pros and cons of using Google Scholar become transparent for the reader.

**Response** >

* Discuss the pros and cons.
* Freely available compared to WoS (paid services).
* See “Comparing Google Scholar with Web of Science and Scopus” in Waltman 2016.

**Google Scholar, Microsoft Academic, Scopus, Dimensions, Web of Science, and OpenCitations’ COCI: a multidisciplinary comparison of coverage via citations**

“Unlike WoS and Scopus, Google Scholar follows an inclusive and automated approach, indexing any seemingly academic document that its crawlers can find and access on the web, including those behind paywalls through agreements with their publishers (Van Noorden [2014](https://link.springer.com/article/10.1007/s11192-020-03690-4" \l "ref-CR60" \o "Van Noorden, R. (2014). November 7). Google Scholar pioneer on search engine’s future. Nature..                   https://doi.org/10.1038/nature.2014.16269                                  .)). Additionally, Google Scholar is free to access, allowing users to access a comprehensive and multidisciplinary citation index without charge.”

“Additionally, there is a significant amount of extra coverage in Google Scholar that is not found in any of the other data sources (26% of all citations across all data sources). Google Scholar could therefore make an important contribution to the scientific community by opening its bibliographic and citation data, which would also facilitate the identification of errors such as coverage fluctuations.”

“the final decision about which source to use may depend on properties of the sources other than coverage, such as metadata quality and bulk access options. If these factors are not of overriding importance, however, then Google Scholar is the best choice in almost all subject areas for those needing the most comprehensive citation counts but not needing complete lists of citing sources.”  
  
**Google Scholar, Web of Science, and Scopus: A systematic comparison of citations in 252 subject categories**

“This study provides evidence that GS finds significantly more citations than the WoS Core Collection and Scopus across all subject areas. ”

“In conclusion, this study gives the first systematic evidence to confirm prior speculation ([Harzing, 2013](https://www.sciencedirect.com/science/article/pii/S1751157718303249" \l "bib0085); [Martín-Martín et al., 2018](https://www.sciencedirect.com/science/article/pii/S1751157718303249" \l "bib0145); [Mingers & Lipitakis, 2010](https://www.sciencedirect.com/science/article/pii/S1751157718303249" \l "bib0160); [Prins et al., 2016](https://www.sciencedirect.com/science/article/pii/S1751157718303249" \l "bib0200)) that citation data in GS has reached a high level of comprehensiveness, because the gaps of coverage in GS found by the earliest studies that analysed GS data have now been filled. It surpasses WoS and Scopus numerically in all areas of research, and is greatly superior in the areas where WoS and Scopus have a poor coverage, including the Social Sciences and Humanities. However, at this point there is no reliable and [scalable method](https://www.sciencedirect.com/topics/computer-science/scalable-method" \o "Learn more about scalable method from ScienceDirect's AI-generated Topic Pages) to extract data from GS, and the metadata offered by the platform is still very limited,”

**A review of the literature on citation impact indicators**

“Google Scholar was also launched in 2004. It indexes scholarly literature that is available online on the web. This includes not only publications in journals and conference proceedings, but also for instance books, theses, preprints, and technical reports. Google Scholar is made freely available by Google. It should be emphasized that Google Scholar is of a very different nature than WoS and Scopus. It is primarily a search engine for scholarly literature, and it provides only very limited bibliographic meta data on publications.”

“A general impression obtained from the literature is that Google Scholar suffers from a lack of quality control. Many inaccuracies in Google Scholar are reported in the literature. [Jacsó, 2005](https://www.sciencedirect.com/science/article/pii/S1751157715300900" \l "bib0770), [Jacsó, 2006](https://www.sciencedirect.com/science/article/pii/S1751157715300900" \l "bib0775), [Jacsó, 2010](https://www.sciencedirect.com/science/article/pii/S1751157715300900" \l "bib0780) for instance discusses problems related to content gaps, incorrect citation counts, and phantom data. The possibility of manipulating citation counts in Google Scholar is discussed by [Beel and Gipp (2010)](https://www.sciencedirect.com/science/article/pii/S1751157715300900" \l "bib0155), [Labbé (2010)](https://www.sciencedirect.com/science/article/pii/S1751157715300900" \l "bib0850), and [López-Cózar, Robinson-García, and Torres-Salinas (2014)](https://www.sciencedirect.com/science/article/pii/S1751157715300900" \l "bib0305). Google Scholar is also criticized for its lack of transparency (e.g., [Jacsó, 2005](https://www.sciencedirect.com/science/article/pii/S1751157715300900" \l "bib0770), [Wouters and Costas, 2012](https://www.sciencedirect.com/science/article/pii/S1751157715300900" \l "bib1690)). It is unclear what is covered by Google Scholar and what is not.”

Statistical analyses

**Comment 9** > CV data usually comes with a lot of missing values. However, I cannot find anything about missing values in the text. Is there no missing values in the data? If there are missing values a wonder how have the authors handled the missingness.

**Response** > Thanks for bringing up this issue. Yes, for the 145 PIs in our final dataset, some of them had missing values for certain response variables, and only those with full information were used in each respective model. For example, a PI who is currently an associated professor will have performance and duration data for recruitment but not promotion. This PI will be included in the analyses of recruitment (Model 1, 3, and 5 in Table 1) but excluded in the analyses of promotion (Model 2, 4, and 6 in Table 1). We have added a column to Table 1 denoting the actual sample size (*n*) in each model.

**Comment 10** > Regarding the LMMs the authors are referencing R-packages which is fine. But I recommend the authors to also provide relevant references for the actual methods they use.

**Response** > We have added a citation for GLMM/LMM to the statistical analyses (Line 210-211):

*“we fit generalized linear mixed models (GLMMs) (Bolker et al. 2009) with …”*

**Reference:**

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., & White, J. S. S. (2009). Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in ecology & evolution*, 24, 127-135.

**Comment 11** > What do the authors mean with the following sentence: "Non-significant interactions (p > 0.05) were dropped from our final model results. Did the authors first try all possible interactions for each model and then in the final models they only included the significant interactions. Or did the authors include all interactions in the final model but only show the significant interaction in the results (i.e., Table 1)?

**Response** >

In our original analyses, we first tested all interactions, dropped the non-significant ones, and refit the final model using only the significant interaction terms. So the results in the original Table 1 represented all the terms (main effects and interactions) in the models. However, we have now re-analyzed our data and included only the main effects of predictors in the models because most models do not have significant interactions, and focusing on the main effects can facilitate the interpretation of the results as well as the comparisons between different models.

**Comment 12** > The authors write that they log-transformed the dependent variables "to meet the assumption of normality". (page 9, row190-191). Did the authors test the assumption of normality on the transformed variables?

**Response** >

We have now re-analyzed the data in our revision. For the academic performance and career duration models (Model 1-4), we did not log-transform the responses but fit GLMMS with a Poisson error distribution and a log link function, as the h-index and year duration are all non-negative integers. For the difference in performance models (Model 5 and 6), we fit LMMs with a Gaussian error distribution and an identity link function, as the difference can take positive, zero, or negative values.

The statistical analyses in the results section were updated (Line 208-244):

*“(1) Academic performance before recruitment/promotion (Model 1 and 2). To examine how various factors affect the academic performance before recruitment as a new PI and promotion to full professor, we fit generalized linear mixed models (GLMMs) (Bolker et al. 2009) with the “Before” h-index for recruitment/promotion as the response, year of recruitment/promotion, PhD university origin (Taiwan vs. Foreign), PhD university ranking, and gender (Male vs. Female) as fixed effects, and the institute (department) nested within university as random effects. The GLMMs were fitted with a Poisson error distribution and a log link function as the response is non-negative integers.*

*(2) Duration before recruitment/promotion (Model 3 and 4). To examine how various factors affect duration before recruitment and promotion, we fit GLMMs with the duration before recruitment/promotion as the response, the “Before” h-index for recruitment/promotion, year of recruitment/promotion, PhD university origin (Taiwan vs. Foreign), PhD university ranking, and gender (Male vs. Female) as fixed effects, and the institute (department) nested within university as random effects. The GLMMs were fitted with a Poisson error distribution and a log link function as the response is non-negative integers.*

*(3) Difference in academic performance before and after recruitment/promotion (Model 5 and 6). To examine how various factors affect the difference in academic performance before and after recruitment/promotion, we fit linear mixed-effects models (LMMs) (Bolker et al. 2009) with the difference between “After” and “Before” h-index for recruitment/promotion (“After” h-index minus “Before” h-index) as the response, year of recruitment/promotion, PhD university origin (Taiwan vs. Foreign), PhD university ranking, and gender (Male vs. Female) as fixed effects, and the institute (department) nested within university as random effects. The LMMs were fitted with a Gaussian error distribution and an identity link function.*

*A total of six models (four GLMMs and two LMMs) were performed using the glmer()/lmer() function in the “lme4” package (Bates et al. 2015). Only full observations were used in each model (observations with any missing entry were omitted; see Table 1 for the actual sample size for each model). The assumption of equal variance and normality were assessed using residual plots and QQ-plots. Significance (α = 0.05) of model coefficients were tested (Wald chi-square test) using the Anova() function in the “car” package (Fox and Weisberg 2019). All analyses were performed in R version 4.2.2 (R Development Core Team 2022).”*

For each model, the assumption of equal variance and normality were both checked using a residual plot and a QQ-plot (the below plots were generated using the R package “performance”):

|  |
| --- |
| Model 1  Performance_recruitment |
| Model 2  Performance_promotion |
| Model 3  Duration_recruitment |
| Model 4  Duration_promotion |
| Model 5  Diff_recruitment |
| Model 6  Diff_promotion |

Overall, there is no severe violation of the assumptions. The residuals in Model 5 seem to deviate from the line at the top-right corner. Nonetheless, this should not be a major issue as studies have shown that regression models are fairly robust to moderate degree of non-normality (Knief and Forstmeier 2021; Schielzeth et al. 2020).

**References:**

Knief, U. & Forstmeier, W. (2021). Violating the normality assumption may be the lesser of two evils. *Behavior Research Methods*, 53, 2576-2590.

Schielzeth, H., Dingemanse, N.J., Nakagawa, S., Westneat, D.F., Allegue, H., Teplitsky, C., Réale, D., Dochtermann, N.A., Garamszegi, L.Z. & Araya‐Ajoy, Y.G. (2020). Robustness of linear mixed‐effects models to violations of distributional assumptions. *Methods in ecology and evolution*, 11, 1141-1152.

*Results*

**Comment 13** > The authors should provide descriptive statistics for their data and variables. Either in the results section or in the Materials and Methods section. This is important so that the reader can get an overview of the data and its properties.

**Response** > We did provide an overview of our data in the first paragraph of the original manuscript (now at Line 247-254 in the revised manuscript).

**Comment 14** > The authors should be clearer about how they use p-values and how they interpret them, especially since their data is not a random sample. For example, the authors write that "PhD university origin, ranking, and gender had no effect on the duration either before recruitment or before promotion (page 10, row 49-54). In Table 1 I can observe that, e.g., the coefficient for the "PhD university origin" is 3.48 which indicates a positive effect and that the p-value is 0.06. I take it that the chosen significance level in the manuscript is 0.05. 0.06 is not that much higher than 0.05. From my perspective I would say that there is a positive effect but that the p-value indicates that there are some uncertainty and that interpretation should therefore be done with some caution. There are two issues here. First, how does the authors define and use p-values in the manuscript? This is not clear. Second, the sample is not a random sample so it seems a bit strange to be super strict about the p-values and e.g., conclude that there is an effect if the p-value is 0.04999 and conclude that there is no effect of the p-value is 0.05, regardless of the size of the coefficient, taking sample size into consideration, etc. The use of p-values does in either case not live up to the required assumptions for making real inference due to the non-random sample. Overall (i.e., this is a recommendation for all the results and not just for the example I provided regarding the "PhD university origin"). The sample size is small (N = 145) and a larger sample size would likely produce significant results.  My suggestion for the authors is to adopt a less dichotomous and more nuanced strategy for interpreting the results of their analyses, e.g., using confidence intervals for determining uncertainty together with the p-values, and not dogmatically approve or disprove of an effect just by looking at the p-values.

**Response** > Thanks for bringing up the issue of interpreting the results based on *p*-values and the suggestion for it. Yes, our significance level was set at 0.05 (Line 242). To avoid the arbitrary cutoff at *p* = 0.05, for those regression coefficients with *p*-values between 0.05 and 0.10, instead of simply stating no effects, we now describe the predictor effects in our results and use “tended to” or “although not statistically significant” to indicate that the *p*-values are not below 0.05 yet there is still some statistical support for the effects. These include:

* *“PIs with Taiwanese PhD degrees tended to have longer durations before recruitment”* (Line 270-271) [*p* = 0.08]
* *“The difference in performance before and after promotion to full professor, although not statistically significant, also decreased over years”* (Line 282-284) [*p* = 0.09]
* *“the difference tended to be higher for PIs with foreign degrees compared to those with Taiwanese degrees”* (Line 284-285) [*p* = 0.06]

Additionally, we have added the 95% confidence limits to Table 1 to show the uncertainty around the regression estimates (also see our response to Comment 17).

**Comment 15** > Can the authors complement the analyses with effect sizes so that it become easier for the reader to understand the size of the effects?

**Response** > Thanks for the suggestion. We have added the regression coefficients (*β*) of the predictors to Table 1.

**Comment 16** > Can the authors transform back the coefficients so that it becomes easier to interpret the actual effects.

**Response** > We have re-analyzed our data and no response was log-transformed in any of the models (also see our response to Comment 10 for more details regarding model fitting).

**Comment 17** > I believe that there is to little information in Table 1 regarding the models and outcomes. As a reader it is difficult to properly assess the results of the analysis. Standard errors and confidence intervals should be included. Some kind of model of fit measure should be included.

**Response** > Thanks for the suggestion. We have added the standard errors and 95% confidence limits of the regression coefficients to Table 1.

**Comment 18** > As I understand it Table 1 present 6 models. This should be more clearly presented in the table, i.e., that each dependent variable in the first column denotes a specific model.

**Response** > Thanks for the suggestion. Yes, there were a total of 6 models in our analyses. We have added a first column to Table 1 denoting each model.

**References provided by Reviewer 2**

Acuna, D. E., Allesina, S., & Kording, K. P. (2012). Predicting scientific success: Daniel E. Acuna, Stefano Allesina and Konrad P. Kording present a formula to estimate the future h-index of life scientists. Nature, 489(7415), 201-202. [https://doi.org/10.1038/489201a](https://doi.org/10.1038/489201a" \t "_blank)

Alonso, S., Cabrerizo, F. J., Herrera-Viedma, E., & Herrera, F. (2009). h-Index: A review focused in its variants, computation and standardization for different scientific fields. Journal of informetrics, 3(4), 273-289.

Bornmann, L., & Daniel, H. D. (2007). What do we know about the h index?. Journal of the American Society for Information Science and technology, 58(9), 1381-1385.

Bornmann, L., & Daniel, H. D. (2009). The state of h index research: Is the h index the ideal way to measure research performance?. EMBO reports, 10(1), 2-6.

Bornmann, L., & Williams, R. (2017). Can the journal impact factor be used as a criterion for the selection of junior researchers? A large-scale empirical study based on ResearcherID data. Journal of Informetrics, 11(3), 788-799. [https://doi.org/10.1016/j.joi.2017.06.001](https://doi.org/10.1016/j.joi.2017.06.001" \t "_blank)

Danell, R. (2011). Can the Quality of Scientific Work Be Predicted Using Information on the Author's Track Record? Journal of the American Society for Information Science and Technology, 62(1), 50-60. [https://doi.org/10.1002/asi.21454](https://doi.org/10.1002/asi.21454" \t "_blank)

Havemann, F., & Larsen, B. (2015). Bibliometric indicators of young authors in astrophysics: Can later stars be predicted? Scientometrics, 102(2), 1413-1434. [https://doi.org/10.1007/s11192-014-1476-3](https://doi.org/10.1007/s11192-014-1476-3" \t "_blank)

Hirsch, J. E. (2007). Does the h index have predictive power? Proceedings of the National Academy of Sciences, 104(49), 19193-19198. [https://doi.org/10.1073/pnas.0707962104](https://doi.org/10.1073/pnas.0707962104" \t "_blank)

Lindahl, J. (2018). Predicting research excellence at the individual level: The importance of publication rate, top journal publications, and top 10% publications in the case of early career mathematicians. Journal of Informetrics, 12(2), 518-533.

Waltman, L., & Van Eck, N. J. (2012). The inconsistency of the h‐index. Journal of the American Society for Information Science and Technology, 63(2), 406-415.