



Can the presence of an author photograph and biography have an impact on article citations? The case of chemistry and chemical engineering

Tahereh Dehdarirad 

Department of Communication and Learning in Science, Chalmers University of Technology, Gothenburg, Sweden

an open access  journal

Citation: Dehdarirad, T. (2022). Can the presence of an author photograph and biography have an impact on article citations? The case of chemistry and chemical engineering. *Quantitative Science Studies*, 3(4), 1024–1039. https://doi.org/10.1162/qss_a_00219

DOI:
https://doi.org/10.1162/qss_a_00219

Peer Review:
https://publons.com/publon/10.1162/qss_a_00219

Supporting Information:
https://doi.org/10.1162/qss_a_00219

Received: 19 March 2022
Accepted: 17 September 2022

Corresponding Author:
Tahereh Dehdarirad
Tahereh@chalmers.se

Handling Editor:
Ludo Waltman

Copyright: © 2022 Tahereh Dehdarirad.
Published under a Creative Commons
Attribution 4.0 International (CC BY 4.0)
license.



Keywords: author characteristics, citation practices, multilevel regression analysis, photograph and biography, random effect within-between (REWB) models, repeated measures

ABSTRACT

The aim of this study was to investigate whether the presence of an author photograph and biography in scientific articles could have an impact on article citations. The impact of a photograph and biography, in combination with certain author characteristics (i.e., gender, affiliation country (measured as whether the author was affiliated with a high-income country or not), and scientific impact (measured as whether the author was a high-impact author or not)), was also examined, while controlling for several covariates. This study focused on a sample of articles published in the time span of 2016–2018 in chemistry and chemical engineering journals by Elsevier. The articles were downloaded from Scopus. The analysis was done using random effects within-between model analyses. Within authors, the results showed no significant impact of author photograph and biography on citations. Different patterns were found for visibility of articles when the presence of an author photograph and biography was combined with author characteristics. While being affiliated to a high-income country and being a high-impact author had a positive impact on citations, gender (female) had a negative impact. For gender, there was a small citation disadvantage of 5% for female authors when they provided a photograph and biography.

1. INTRODUCTION

The process of scientific citations and visibility in the scientific community seems to be subject to implicit biases (Dion & Mitchell, 2020). Previous research has shown that author characteristics such as academic age (Mahbuba & Rousseau, 2011) or the prestige of affiliated country or institutions (Didegah, Bowman, & Holmberg, 2018) might affect the visibility of scientific works in terms of citations. For example, a scientific work published by a famous person (Mahbuba & Rousseau, 2011), or a country or an institute with high prestige in terms of citation (Didegah et al., 2018) might have higher visibility in terms of citations. The result of a more recent study by Nielsen and Andersen (2021) showed that the top 1% most-cited scientists, who are often affiliated with high-ranking institutions in the Western world, receive more citations.

The Matthew effect has been applied to explain these inequalities and systematic deviations in the number of citations from that of expected citations (Vinkler, 2010). The Matthew effect refers to “greater increments of recognition for particular scientific contributions to scientists of considerable repute and the withholding of such recognition from scientists who have not yet made their mark” (Merton, 1973).

In addition to discrimination based on the abovementioned variables, reduced recognition might arise from other factors, such as gender (Mahbuba & Rousseau, 2011). According to a League of European Research Universities report by Gvozdanović and Maes (2018), implicit bias has been a significant impediment to women's advancement in academic careers. Implicit biases might affect citation practices in a way that make men's research be viewed as more central or important in a scientific field (the Matthew effect), whereas women's work would be underrecognized (the Matilda effect) (Ghiasi, Mongeon et al., 2018; Dion, Sumner, & Mitchell, 2018; Dion & Mitchell, 2020). Previous research has shown that in disciplines such as social and political sciences where there are more female scholars, women receive a similar or higher number of citations. This might mean a reduction of the Matthew effect. However, gender citation gaps may persist if the Matilda effect and implicit biases in citation practices occur (Dion et al., 2018).

Similarly, publications from countries with higher research and development investment (Mahbuba & Rousseau, 2011) might receive more citations or attention on social media. Harris, Macinko et al. (2017) used the Implicit Association Test with healthcare professionals to study the association between good research and richness of countries (measured as their gross domestic product per capita [GDP]). The results showed that the majority of participants associated good research with rich countries. Implicit associations such as these might disfavor research from lower income countries in research evaluation and citation practices, and consequently their visibility (Harris et al., 2017). Mishra and Wang (2021) studied 55 million publications covering STEM research from 218 countries. The results showed that while there is a clear convergence (in terms of output and impact) among the high-income and upper middle income countries, there is a widening gap that segregates the lower middle and low-income regions from the higher income regions.

The presence of an author photograph and biography in scientific articles is another characteristic that could potentially create bias in citation practices and sharing or engaging with articles on social media platforms. Previous research has shown that the way researchers present themselves through their profile image or biography could have an impact on credibility and trust formation (Francke, 2021). However, the way authors are perceived through their photograph or biography could be affected by author characteristics such as gender, country, or scientific impact, which have been previously shown to create bias in citations. In fact, appearance has been shown to have a different impact on the perceived likelihood of being a scientist for male and female academics (Banchefsky, Westfall et al., 2016).

While many studies have examined factors causing bias in scientific practices, and more generally the driving factors of citations (Didegah et al., 2018), the potential impact of authors' photographs and biographies on citation has remained a not very well explored research question. Fidrmuc and Paphawasit (2018) studied the impact of physical attractiveness on the productivity of authors in the economics field. While the results showed a significantly positive effect of authors' attractiveness on both journal quality and citations, the impact on citations disappeared after controlling for journal quality. The result of a study by Dehdarirad and Sotudeh (2017) in chemistry showed a very small citation superiority in favor of women in papers where authors included their photograph and biography. This study aims to extend the latter line of research by investigating whether, and, if so, how, the presence of author photographs and biographies in scientific articles can have an impact on the visibility of articles in terms of citation. To ensure that author photographs and biographies were presented in the same way, this study focused on a sample of articles in chemistry and chemical engineering journals published by Elsevier.

To achieve the aim, the following research question is addressed:

- Is there any difference in log transformed citations received by papers based on gender, country, and impact of authors when they shared their photograph and biography, in comparison with when they did not?

While addressing the question, the impact of other covariates that might influence the number of citations was controlled (see Section 2.2).

2. METHODOLOGY

2.1. Data Collection

The data in this study is based on articles in 121 chemistry and chemical engineering journals published by Elsevier, which are indexed in the Scopus database. The data were retrieved from Scopus in February 2021. The studied time period was 2016–2018.

By consulting the author information pack of each journal, the journal articles were divided into two groups:

- Set 1: Authors were asked to provide a photograph and biography.
- Set 2: Authors were not asked to provide a photograph and biography.

As, in some cases, the authors might not provide their photograph, all articles in set 1 were double-checked for the existence of photographs using the ScienceDirect Elsevier Object Retrieval API (https://dev.elsevier.com/sciencedirect.html#/Object_Retrieval).

In the next step, articles in set 1 were paired with articles in set 2, based on a common author. In other words, the articles in both sets were written by a common author under two different conditions. In set 1, authors provided a biography and photograph, whereas in set 2, authors did not. As a result, the data for this study comprised two sets. Set 1 comprised 4,572 articles, whereas set 2 comprised 9,292 articles. This accounted for 6,284 authors and 32,005 observations, which were used in citation analysis (Models 0 and 1). Author disambiguation was mainly done using a combination of Scopus authors IDs and ORCID, where both were available. If ORCID was not available, then Scopus author ID was used. In the few cases where the name of an author appeared under two author IDs, a manual check was done by searching online for authors' web pages, where they provided a list of their publications. This was done to check whether the publications belonged to the same author or not.

The common authors were either first or second authors on a paper. The reason for selecting these authors was their higher visibility in a paper where authors provided their photograph and biography. Additionally, this made it possible to control for the impact of author byline order in both sets.

2.2. Covariates

In this section, details regarding data collection and processing of the covariates used in the regression analyses (see Section 2.3) are provided. Reasons are also provided why they were included as covariates in the regression models.

Several of the covariates included in the regression models capture different aspects that might have an impact on the number of citations.

Collaboration, measured as number of authors and number of countries, and journal impact, measured as source normalized impact per paper (SNIP), have been among the most important factors associated with citations (Didegah & Thelwall, 2013; Didegah et al., 2018).

As reputation can be considered as an important signal of trustworthiness and quality (Petersen, Fortunato et al., 2014), it is expected that papers by well-reputed authors or from rich countries (high income) will receive more citations. Previous research has shown a relationship between authors' and countries' prestige and the number of citations (Didegah et al., 2018; Dehdarirad & Karlsson, 2021). Gender has also been shown to affect citation counts in the field of chemistry (Day, Corbett, & Boyle, 2020). Thus, in this paper, these author characteristics for common authors (gender, scientific age, and GDP level of affiliation country) were controlled. Additionally, the impact of the rest of the authors (coauthors of the common author) on each paper was controlled in terms of these characteristics.

To detect the gender of authors, a combination of gender API (<https://gender-api.com/>) and manual checking was carried out. First, gender API was used to conduct a search using the first name of authors. Then, in cases of gender-neutral, unknown, initials, or where the accuracy was lower than 80%, the names were checked manually using internet searches. For common authors, gender was treated as a binary variable with male (0) and female (1). For the other authors, the female proportion per paper was calculated. In the regression models this variable is named "female proportion."

The scientific age of authors was calculated by using the geometric mean of citations for both common (named "Author impact") and other authors (named "Rest of authors' impact"). The number or log-transformed number of publications and citations of an author has been previously defined as the professional or scientific age of an author (Mishra, Fegley et al., 2018; Andersen, Schneider et al., 2019). For this paper, the authors were divided into quartiles based on their geometric mean of citations. Then, a dummy variable was created which showed whether the author belonged to the top high-impact quartile (1) or not (0).

GDP data were gathered from the World Data Bank website (<https://data.worldbank.org/indicator/NY.GDP.MKTP.CD>). The World Bank assigns the world's economies into four income groups—high, upper middle, lower middle, and low. For common authors, a dummy variable was created that indicated whether the studied author was affiliated to a high-income country (1) or not (0). This is named *GDP level* in the regression models. To control for investment in GDP for the other authors on a paper, the GDP values for their affiliated countries during 2016–2018 were obtained and an average for this time period was calculated. Then for each paper, GDP average values for their corresponding countries were averaged. This is named *Rest of authors' GDP* in the regression models.

Open access has been shown to be an important factor for citations (Eysenbach, 2006; Hajjem, Harnad, & Gingras, 2006; Gargouri, Hajjem et al., 2010). In this study, open access was treated as a binary variable with OA (1) and non-OA (0).

As article topic has been shown to influence citation counts (Gallivan, 2012; Wang, Jiao et al., 2020), a categorical variable was created which grouped articles into five broad headings: "Biochemistry," "Applied," "Physical, Inorganic, and Analytical," "Organic," and "Macromolecular." The Chemical Abstracts Service headings were used to tag the articles.

Regarding abstract readability, the Flesch Reading Ease Score was used as it is the most used measure of text readability and has been used in other studies (Didegah et al., 2018). The R *quanteda* package (Benoit, Watanabe et al., 2018) was used to calculate this score for each abstract. The highest possible score is 121.22, and there is no lower limit. The higher the score, the easier the text is to understand.

Finally, in the regression models, time since publication was considered as an offset variable.

2.3. Data Analysis

2.3.1. Models 1 and 2: Mixed and hybrid (within-between) models

In both models in this section, the outcome variable ($\ln(\text{citation} + 1)$) is a continuous variable and there is a random effect associated with the intercepts for each author (Author_ID). Authors are also nested within countries (Country/Author_ID). Furthermore, there are several fixed effects predictors with fixed slopes (see Table 1).

There are two types of predictors (time-varying and time-invariant) in the models. In the context of applying multilevel models to repeated measure (longitudinal) data problems, time-varying predictors will appear at level 1, because they are associated with specific measurements, whereas time-invariant predictors will appear at level 2 or higher, because they are associated with the individual (or a higher data level) across all measurement conditions (Finch, Bolin, & Kelley, 2019). Thus, in this study, photograph and biography, GDP for rest of authors, open access, readability, rest of authors' impact, female proportion, number of countries, number of authors, and topic are level 1 predictors. Gender, Journal, author impact, and GDP level are level 2 predictors.

First a linear mixed model (LMM) using the *lme4* R package (Bates, Mächler et al., 2015) was applied to the $\ln(\text{citation} + 1)$. Citation distributions follow close to a discretized lognormal distribution. Thus, for regression analyses, the best option is to use ordinary least squares regression applied to the natural logarithm of citation counts plus one (Thelwall, 2016). The log transformation prevents individual highly cited articles from dominating the results (Thelwall & Sud, 2020). To compare all models and to decide which variables to include in the models, the *anova()* function was used. It was also checked to whether keep the random effects for authors and journals in the models or not. To do this, for both authors and journals, a linear model (LM) without random effect was fitted to the data using the *lm* function. Then, LM and LMM models were compared in terms of their BIC (Bayesian information criterion) and

Table 1. Variables by type

Variable type	Variable
Random factor	Country/Author_ID
Fixed factors	Gender, female proportion, GDP level of affiliation country, rest of authors' GDP, author impact, rest of authors' impact, open access, Journal, readability, number of countries, number of authors, topic
Repeated measures	Photograph and biography
Dependent variable	$\ln(\text{citation} + 1)$

Note. In the models, "rest of authors' impact" was not entered as it did not improve the models.

RMSE (root mean squared error) using the *Compare_performance* command in the R *performance* package (Lüdtke, Ben-Shachar et al., 2021). For authors, the results showed that the LMM model had lower values of BIC and RMSE and a 100% performance score in comparison to the LM model. The result of a likelihood ratio test for authors using the *rand* function in the *lmerTest* package (Kuznetsova, Brockhoff, & Christensen, 2017) also showed a significant *p*-value ($p < 0.001$). Thus, the random effect for authors was retained in the model. For journals, the results showed no difference between the LMM and LM models. Thus, the simpler model without the random effect for journals was preferred. Thus, journal was entered as a factor variable in the models. The inclusion of journal fixed effects accounted for factors that do not vary within a journal, such as its specialty or impact, that might be correlated with citations.

Of the models fitted, the mixed model with the command below was the most parsimonious for the data.

```
Model 1=lmer (Log_normalized_citations ~ photo&bio + offset(log(time)) + author_gender*  
photo&bio + female_proportion + GDP_level + GDP_rest_of_authors + OA + author_impac*  
photo&bio + GDP*photo&bio + number_of_authors + Topic + Factor(Journal) + readability +  
number_of_countries + (1 | country/AU_ID), data=citation_data, REML=FALSE, control=  
lmerControl(optimizer="optimx",optCtrl=list(method='nlminb'))
```

The R *performance* package was used to check the model fit in terms of outliers and influential observations, normality of residuals for both fixed and random effects, collinearity, and heteroscedasticity (see Supplementary material, Section 1).

LMMs are statistical models for continuous outcome variables, in which the residuals are normally distributed but may not be independent or have constant variance. Studies with clustered, longitudinal, or repeated measures data may be appropriately analyzed using LMMs (West, Welch, & Galecki, 2015). The case in this study involves repeated measures data, in which multiple measurements regarding the citations were made on the same subject (authors) under two different conditions (with and without photograph and biography). Repeated measures data sets can be a type of two-level data, in which Level 2 represents the subjects (authors in this case) and Level 1 represents the repeated measurements made on each subject (observations) (see Table 1). Covariates measured at Level 2 of the data describe between-subject variation and Level 1 covariates describe within-subject variables (West et al., 2015). These models are called *mixed effects*, as they combine both fixed effects and random effects. Random effects are estimated with shrinkage (partial pooling), while fixed effects are estimated using least squares (or, more generally, maximum likelihood) (no pooling). No pooling assumes that the residual variance is the same within each group (Gelman & Hill, 2008) whereas partial pooling estimates combine within and between-group variance.

After fitting a mixed effect model (Model 1), it was investigated whether within and between effects were equal. Mixed effect models assume that between and within effects are equal (Bell, Fairbrother, & Jones, 2019, p. 1057). In mixed effect models, covariates at different levels can be related to outcome variables in quite different ways (Bell et al., 2019). When within and between are not equal, the model will produce a weighted average of the two, which will have little substantive meaning (Raudenbush & Bryk, 2002, p. 138; Bell et al., 2019). Additionally, in the latter case, combining these two relationships inevitably leads to biased estimates of both, as each will be drawn towards the other (Bell, Jones, & Fairbrother, 2018). This assumption was checked by testing the equality of the coefficients in a random

effect within-between model (REWB) (Model 2). As there were differences between the terms, the REWB model was used. Additionally, the performance of Models 1 and 2 was compared in terms of BIC and RMSE. As Model 2 showed 100% better performance, the results of the REWB model (Model 2) were reported in the paper. Model 2 was fitted with the command below:

Model 2=lmer (Log_normalized_citations ~ photo&bio (within) + photo&bio (between) + offset(log(time)) + within (product of author_gender and photo&bio) + between (product of author_gender and photo&bio) + female_proportion + GDP level + within (GDP_rest_of_authors) + between (GDP_rest_of_authors) + within (OA) + between (OA) + within (product of photo&bio and author impact) + between (product of photo&bio and author impact) + within (product of photo&bio and GDP) + between (product of photo&bio and GDP) + within (number_of_authors) + between (number_of_authors) + Topic + Factor(Journal) + within (readability) + between(readability) + within (number_of_countries) + between (number_of_countries) + (1 | country/AU_ID), data=citation_data, REML= FALSE, control= lmerControl(optimizer="optimx",optCtrl=list(method='nlminb'))

See Supplementary material, Section 2 for diagnostics of Model 2.

$$\text{Log-normalized-citation}_i \sim N(\mu, \sigma^2)$$

$$\mu = \alpha_{j[i],k[i]} + \beta_1(x1) + \beta_2(x2) + \beta_3(x3) + \beta_4(x4) + \beta_5(x5) + \beta_6(x6) + \beta_7(x7) + \beta_8(x8) + \beta_9(x9) + \beta_{10}(x10) + \beta_{11}(x11) + \beta_{12}(x12) + \beta_{13}(x13) + \beta_{14}(x14) + \beta_{15-134}(as.factor(x15))$$

$$\alpha_j \sim N(\gamma_0^\alpha + \gamma_1^\alpha(x16) + \gamma_2^\alpha(x17) + \gamma_3^\alpha(x18) + \gamma_4^\alpha(x19) + \gamma_5^\alpha(x20) + \gamma_6^\alpha(x21) + \gamma_7^\alpha(x22) + \gamma_8^\alpha(x23) + \gamma_9^\alpha(x24) + \gamma_{10}^\alpha(x25) + \gamma_{11}^\alpha(x26) + \gamma_{12}^\alpha(x27), \sigma_{\alpha_j}^2), \text{ for authors : countries } j = 1, \dots, J$$

$$\alpha_k \sim N(\gamma_0^\alpha + \gamma_1^\alpha(x28), \sigma_{\alpha_k}^2), \text{ for number of countries } k = 1, \dots, K$$

where:

- X1: Within_Photo&Bio
- X2: Within (Photo&Bio and gender)
- X3: Within-Female-proportion
- X4: Within (Photo &Bio and High-level income country)
- X5: Within (Photo&Bio and High-impact-author)
- X6: Within_Number-of-authors
- X7: Within_Number-of-countries
- X8: Within_Rest-of-authors-GDP
- X9: Within_Open-Access
- X10: Within_Readability
- X11–X14 Topic- Applied; Topic- Physical, inorganic, and analytical; Topic- Organic; Topic- Macromolecular
- X15: Journal
- X16: Between_Photo &Bio
- X17: Between (Photo&Bio and gender)
- X18: Gender

- X19: Between_Female-proportion
- X20: Between (Photo&Bio and High-level income country)
- X21: Between (Photo&Bio and High-impact-author)
- X22: Between_Number-of-authors
- X23: Between_Number-of-countries
- X24: Between_Rest-of-authors-GDP
- X25: Between_Open-Access
- X26: Between_Readability
- X27: High-impact-author
- X28: High-level income country

REWB models (also known as *hybrid models*) and correlated random-effects models (CRE) are flexible modeling specifications that separate within and between-cluster effects and allow for both consistent estimation of Level 1 effects and inclusion of Level 2 variables (Schunck, 2013). The between effect in these models cannot indicate causality because it only represents the mean difference (between groups) over all measurement points. In hybrid models, similar to FE (fixed effect) models, a within effect estimator only uses the within variance. FE (within) models are not generally suited for estimating absolute group differences. However, using interaction between group dummies (e.g., gender) and time-varying variables (picture and photograph) in FE models, it was possible to estimate the differences in the coefficients of covariates by group (Collischon & Eberl, 2020). This methodology has also been applied in econometrics studies such as Wooldridge (2010). Additionally, hybrid models allow for the possibility of including a random intercept and random slopes to control for random effects and thus allow the Level 1 (within) estimators to vary between individuals (Schumann & Kuchinke, 2020).

To estimate between and within effects in one model, the cluster-specific mean for time varying variables was calculated (between values) first. The second step was to create the deviation scores, which is also known as *group mean centering* (within values) (Schunck, 2013). The final step was regarding interaction terms. In this paper, there are three interaction terms for the time-varying variable (photograph and biography) with time-invariant variables (gender, being high-impact author, being affiliated with a high-income country). Following the methodology in Schunck (2013), three new variables were calculated by creating the product of the time-varying variable (photograph and biography) and each of the time-invariant variables. Then, for each of these three variables, their individual cluster mean and deviation scores were calculated and entered in Models 2 and 3.

Finally, Model 3 was fitted using the *marginaleffects* R package (Arel-Bundock, 2022), which shows the within-between model with marginal effects. This was done by restricting the data to those authors whose status changed in terms of providing a photograph and biography. Thus, the authors who always provided a photograph and biography and those that never did were excluded from the analysis. As a result, the number of observations reduced to 29,134, accounting for 4,855 authors.

3. RESULTS

3.1. REWB model: Models 2a, 2b, and 3

In the models, the effect of photograph and biography is separated in both within and between parts. The effect for the other time-varying variables is also separated in within and between

Table 2. Random effects within-between model for comparison of authors with and without photograph and biography

Predictors	Model 2a: REWB model with no covariates		Model2b: REWB model with covariates	
	Estimates	CI	Estimates	CI
(Intercept)	1.42 ^{***}	1.35–1.49	1.71 ^{***}	1.55–1.86
Within (Photo & Bio) [Yes]	0.08 ^{***}	0.07–0.09	–0.31	–1.20–0.58
Between (Photo& Bio) [Yes]	0.07	–0.01–0.14	–0.31	–1.19–0.57
Gender (Female)			0.08 [*]	0.00–0.16
High level Income [Yes]			–0.12 [*]	–0.24 – –0.01
High-Impact Author [Yes]			0.20 ^{***}	0.13–0.27
Within (Female proportion)			0.06 ^{***}	0.03–0.08
Between (Female proportion)			–0.10 [*]	–0.18 – –0.02
Within (Rest of authors' GDP)			–0.18 ^{***}	–0.26 – –0.10
Between (Rest of authors' GDP)			–0.15 [*]	–0.27 – –0.02
Within (Number of authors)			0.45 ^{***}	0.38–0.53
Between (Number of authors)			0.52 ^{***}	0.33–0.71
Within (Number of countries)			0.07	–0.04–0.18
Between (Number of countries)			0.68 ^{***}	0.28–1.07
Within (Open Access) [Yes]			–0.04 ^{***}	–0.06 – –0.02
Between (Open Access) [Yes]			–0.10 ^{***}	–0.16 – –0.05
Within (Readability)			0.01	–0.05–0.08
Between (Readability)			–0.54 ^{***}	–0.74 – –0.33
Topic [Applied]			0.05 ^{***}	0.03–0.07
Topic [Physical, inorganic, and analytical]			0.02	–0.01–0.04
Topic [Organic]			–0.03 [*]	–0.05 – –0.00
Topic [Macromolecular]			–0.12	–0.45–0.22
Within [(Photo & Bio) [Yes] × Gender [female]]			–0.05 ^{***}	–0.07 – –0.03
Between [(Photo & Bio) [Yes] × Gender [female]]			–0.06	–0.20–0.08
Within [(Photo & Bio) [Yes] × High Level Income [Yes]]			0.05 ^{***}	0.03–0.06
Between [(Photo & Bio) [Yes] × High Level Income [Yes]]			–0.01	–0.13–0.12
Within [(Photo & Bio) [Yes] × High-Impact Author]			0.02 [*]	0.00–0.04
Between [(Photo & Bio) [Yes] × High-Impact Author]			0.06	–0.06–0.19

Downloaded from http://direct.mit.edu/qss/article-pdf/3/4/1024/2070739/qss_a_00219.pdf by guest on 13 June 2024

Table 2. (continued)

Predictors	Model 2a: REWB model with no covariates		Model2b: REWB model with covariates	
	Estimates	CI	Estimates	CI
Random effects				
σ^2	0.16		0.12	
τ_{00}	0.46 _{AuthorID:country}		0.30 _{AuthorID:country}	
	0.05 _{Country}		0.02 _{Country}	
N	6,284 _{AuthorID}		6,284 _{AuthorID}	
	87 _{Country}		87 _{Country}	
Observations	32,005		32,005	
Marginal R^2 /Conditional R^2	0.002/0.754		0.269/0.805	

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Note. Ref group for topics was Biochemistry.

parts in Models 2b and 3. In the regression tables, marginal R^2 provides the variance explained only by fixed effects and conditional R^2 provides the variance explained by the entire model (i.e., both fixed effects and random effects). σ^2 is the within-author variance whereas τ_{00} is the *between-author* variance.

Looking at the within coefficient in Model 2a (see Table 2), the impact of photograph and biography was significant. In other words, providing a photograph and biography was associated with an increase by 8% in citations in the model without covariates. However, after controlling for the covariates, the effect became insignificant. As can be seen from Models 2b and 3 (see Table 3), both within and between coefficients for photograph and biography were not significant. Moving from Model 2a to Model 2b, marginal R^2 increased from 0.002 to 0.27. This means that 27% of the variance in Model 2b is explained by the inclusion of fixed effect predictors.

For within coefficients, when looking at the interaction terms of photograph and biography with gender, affiliation to a high-income country, and being a high-impact author, the impact was significant. While both GDP level (being affiliated to a high-income country) and author impact had a positive mediating impact on citations, gender (female) had a negative impact. For gender, there was a small citation disadvantage of 5% for female authors when they provided a photograph and biography.

There was a small citation advantage of 5% for being affiliated with a high-income country when authors provided a photograph and biography, compared to when they did not. For being a high-impact author, the within authors coefficient indicated a very small citation advantage of 2% when authors provided a photograph and biography.

Model 3 shows the within-between model with marginal effects. When comparing the results in Models 3 and 2b, it can be seen that the within coefficients for photograph and biography and its interaction terms remained identical.

Table 3. Marginal effects for Model 3, examining the effects of change in making photograph and biography available

Predictors	Model 3: REWB with marginal effects	
	Estimates	CI
(Intercept)	1.80 ^{***}	1.56–2.05
Within (Photo & Bio) [Yes]	–0.30	–1.17–0.58
Between (Photo & Bio) [Yes]	–0.41	–1.37–0.55
Gender (Female)	0.13	–0.09–0.35
High-Income [Yes]	–0.19	–0.41–0.02
High-Impact Author [Yes]	0.29 ^{**}	0.09–0.50
Within (Female proportion)	0.05 ^{***}	0.03–0.08
Between (Female proportion)	–0.10 [*]	–0.19 – –0.01
Within (Rest of authors' GDP)	–0.20 ^{***}	–0.27 – –0.12
Between (Rest of authors' GDP)	–0.13	–0.28–0.02
Within (Number of authors)	0.49 ^{***}	0.41–0.57
Between (Number of authors)	0.50 ^{***}	0.28–0.71
Within (Number of countries)	0.06	–0.05–0.17
Between (Number of countries)	0.89 ^{***}	0.42–1.36
Within (Open Access) [Yes]	–0.04 ^{***}	–0.06 – –0.02
Between (Open Access) [Yes]	–0.14 ^{***}	–0.21 – –0.07
Within (Readability)	–0.00	–0.07–0.07
Between (Readability)	–0.64 ^{***}	–0.88 – –0.40
Topic [Applied]	0.05 ^{***}	0.02–0.07
Topic [Physical, inorganic, and analytical]	0.01	–0.01–0.04
Topic [Organic]	–0.03 ^{**}	–0.05 – –0.01
Topic [Macromolecular]	–0.12	–0.46–0.21
Within [(Photo & Bio) [Yes] × Gender [female]]	–0.05 ^{***}	–0.07 – –0.03
Between [(Photo & Bio) [Yes] × Gender [female]]	–0.19	–0.66–0.28
Within [(Photo & Bio) [Yes] × High-Income [Yes]]	0.05 ^{***}	0.03–0.06
Between [(Photo & Bio) [Yes] × High-Income [Yes]]	0.14	–0.25–0.54
Within [(Photo & Bio) [Yes] × High-Impact Author]	0.02 [*]	0.00–0.04
Between [(Photo & Bio) [Yes] × High-Impact Author]	–0.18	–0.59–0.24

Table 3. (continued)

Predictors	Model 3: REWB with marginal effects	
	Estimates	CI
Random effects		
σ^2	0.12	
τ_{00} Author_ID:Country	0.30	
τ_{00} Country	0.03	
$N_{\text{authors_ID}}$	4,855	
N_{Country}	86	
Observations	29,134	
Marginal R^2 /Conditional R^2	0.272/0.806	

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

4. DISCUSSION AND CONCLUSION

The aim of this study was to investigate whether, and, if so, how, the presence of an author photograph and biography in scientific articles can have an impact on the visibility of articles in terms of citation. This was done using random effects within-between model analyses. To ensure that authors presented their photographs and biographies in the same way, this study focused on a sample of articles in chemistry and chemical engineering journals published by Elsevier. The impact of photographs and biographies in combination with authors characterizes, for example, gender and affiliation country (GDP level); scientific impact was also examined while controlling for several covariates. The findings regarding these are briefly presented and discussed below.

The results of the within model showed lower citations for papers where authors provided a photograph and biography. However, the effect was insignificant both within and between models when controlling for covariates.

When examining the interaction terms of common authors' characteristics with photograph and biography, the impact of gender (female) was significantly negative. Additionally, the finding showed that when female authors provided a photograph and biography, the estimated citation counts decreased by 5%. This finding is interesting, as chemistry and chemical engineering have been described as examples of relatively high gender-diverse fields (Blosser, 2017; Jarboe, 2019). As indicated in Dion et al.'s (2018) study, while improvements in gender diversity in academia might increase the visibility and impact of scholarly work by women, implicit biases in citation practices might persist. As in this study, the authors provided their photograph and biography in the same way, the finding could be related to implicit bias and how women are perceived and judged by their appearance. The result of this study is to some extent in line with a study by Banchevsky et al. (2016), which suggested that for men there was no impact of gendered appearance on the perceived likelihood of being a scientist, whereas for women their appearance mattered. According to this study, feminine appearance may erroneously signal that women are not well suited for science. As appearance has a powerful and immediate effect on a person's perception of others, variations in facial appearance based on gender could automatically activate gender stereotypes (Banchevsky et al., 2016; Ito & Urland 2003), thereby affecting citation decision-making practice. In social science research, this

could be referred to as an “amplification process,” which allows categorical differences associated with gender to expand in their range of application (Ridgeway, 2011). For example, gendered stereotypes about visual representation of men and women can be transferred to evaluations of the products they produce (articles in this study), with women being disadvantaged (Tak, Correll, & Soule, 2019). The finding is, however, in contrast with Dehdarirad and Sotudeh (2017), who found a very small citation superiority in favor of women in papers where authors provided a photograph and biography. The reason for this contrast could be due to the different methodologies, samples, and time spans used in these studies, as in this paper, by using REWB regression models, it was possible to control for several covariates and estimate Level 1 effects while including Level 2 variables. Additionally, Dehdarirad and Sotudeh (2017) examined a data set of papers published by highly productive authors, whereas the current study did not.

Regarding the interaction terms of photograph and biography with two other studied author characteristics (i.e., being a high-impact author and being affiliated with a high-income country), the results showed a significant positive impact for both these variables. The findings showed small citation advantages of 5% and 2% for being affiliated with a high-income country and being a high-impact author, respectively. The findings seem to suggest that a photograph and biography in combination with author characteristics (being a high-impact author and being affiliated with a high-income country) could have an amplifying bias effect on the number of citations received. This amplifying effect may disfavor research from lower income countries or authors with lower impact to a greater extent in research evaluation and citation practice (Harris et al., 2017). These findings could be explained by a combination of “status characteristics theory” and “amplification process” mechanisms. As mentioned in Nielsen, Baker et al. (2021), a scientist’s country of affiliation could be considered as a “status signal.” Thus, being affiliated with a high-income country might implicitly influence evaluation practices. Additionally, facial appearance is important in the perception of academics and their dissemination of knowledge (Bi, Chan, & Torgler, 2020). Based on the “amplification process” mechanism, these implicit biases about affiliation countries of authors, combined with physical appearance biases, might be transferred to the evaluation of their papers and thereby affect citation practice.

In general, the findings of this study add to the literature on the role that photographs and biographies, in combination with author characteristics, could play in the visibility of articles. The findings are important as they provide insights into how citation practices could be biased by the effect of authors’ photographs and biographies in combination with author characteristics.

Using an REWB model in this paper, it was possible to estimate the effects of time-invariant as well as time-varying predictors in the analysis. It was also possible to include random slopes in the models, thereby allowing the effects of Level 1 variables to vary between clusters. As bibliometric data has a multilevel structure (Mutz & Daniel, 2019), these models could be extended and used in other bibliometrics studies to correct for the dependence of multiple observations per individual. It also allows us to estimate the effect of Level 2 variables while providing effect estimates of Level 1 variables, which are unbiased by a possible correlation with the Level 2 error (Schunck, 2013). The usefulness of statistical analysis of citations using causality in bibliometric studies has also been previously investigated and shown (Bittmann, Tekles, & Bornmann, 2021; Mutz, Wolbring, & Daniel, 2017). As suggested in Bittmann et al. (2021), these techniques could be used as an approach in bibliometrics to estimate effects and remove bias.

In this paper, the visibility of articles in terms of citations was studied from the point of view of photographs and biographies, in combination with author characteristics. However,

engagement with articles might also be dependent on the characteristics of the people who engage with the paper (readers). Thus, in future research, I aim to consider the effect of readers' characteristics too.

This study has some limitations. The findings in this study are limited to a sample of articles in chemistry and chemical engineering. Thus, generalization of the findings to other fields is not advised. From a methodological viewpoint, it is important to remember that despite the advantages of mixed and REWB models, these models cannot control for unmeasured time-varying confounding variables (Bell et al., 2019; Gunasekara, Richardson et al., 2013). However, I attempted to control for several important variables in the models. By doing so, it was possible to increase the possibility of obtaining more precise and reliable results.

ACKNOWLEDGMENT

The author would like to thank Mr. Jonathan Freer at Chalmers University of Technology for proofreading the article.

FUNDING INFORMATION

The author did not receive any funding for this work.

COMPETING INTERESTS

The author has no competing interests.

DATA AVAILABILITY

All relevant data regarding the variables, how they were obtained, R packages, and commands used for regression analysis are within the paper. The diagnostics for regression models are also in the Supplementary material. However, restrictions apply to the availability of the bibliometric data. These data were downloaded under the provision of the institutional standard contract held by the Chalmers University of Technology to Scopus, Elsevier API (https://dev.elsevier.com/sciencedirect.html#/Object_Retrieval), and SciVal API (<https://dev.elsevier.com/scival.html>). I did not have any special access privileges to these databases. Interested researchers may access the Scopus, Elsevier, and SciVal APIs in the same way I did.

REFERENCES

- Andersen, J. P., Schneider, J. W., Jagsi, R., & Nielsen, M. W. (2019). Gender variations in citation distributions in medicine are very small and due to self-citation and journal prestige. *eLife*, 8, e45374. <https://doi.org/10.7554/eLife.45374>, PubMed: 31305239
- Arel-Bundock, V. (2022). *Marginal effects: Marginal effects, marginal means, predictions, and contrasts*. R package version 0.5.0.9000. <https://vincentarelbundock.github.io/marginaleffects/><https://github.com/vincentarelbundock/marginaleffects>
- Banchefsky, S., Westfall, J., Park, B., & Judd, C. M. (2016). But you don't look like a scientist! Women scientists with feminine appearance are deemed less likely to be scientists. *Sex Roles: A Journal of Research*, 75(3–4), 95–109. <https://doi.org/10.1007/s11199-016-0586-1>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Bell, A., Jones, K., & Fairbrother, M. (2018). Understanding and misunderstanding group mean centering: A commentary on Kelley et al.'s dangerous practice. *Quality & Quantity*, 52(5), 2031–2036. <https://doi.org/10.1007/s11135-017-0593-5>, PubMed: 30147154
- Bell, A., Fairbrother, M., & Jones, K. (2019). Fixed and random effects models: Making an informed choice. *Quality & Quantity*, 53(2), 1051–1074. <https://doi.org/10.1007/s11135-018-0802-x>
- Benoit, K., Watanabe, K., Wang, H., Nulty, P., Obeng, A., Müller, S., & Matsuo, A. (2018). quanteda: An R package for the quantitative analysis of textual data. *Journal of Open Source Software*, 3(30), 774. <https://doi.org/10.21105/joss.00774>
- Bi, W., Chan, H. F., & Torgler, B. (2020). “Beauty” premium for social scientists but “unattractiveness” premium for natural scientists in the public speaking market. *Humanities and Social Sciences Communications*, 7(1), 118. <https://doi.org/10.1057/s41599-020-00608-6>
- Bittmann, F., Tekles, A., & Bornmann, L. (2021). Applied usage and performance of statistical matching in bibliometrics: The comparison of milestone and regular papers with multiple measurements of disruptiveness as an empirical example. *Quantitative Science Studies*, 2(4), 1246–1270. https://doi.org/10.1162/qss_a_00158

- Blosser, E. (2017). Gender segregation across engineering majors: How engineering professors understand women's underrepresentation in undergraduate engineering. *Engineering Studies*, 9(1), 24–44. <https://doi.org/10.1080/19378629.2017.1311902>
- Collischon, M., & Eberl, A. (2020). Let's talk about fixed effects: Let's talk about all the good things and the bad things. *KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 72(2), 289–299. <https://doi.org/10.1007/s11577-020-00699-8>
- Day, A. E., Corbett, P., & Boyle, J. (2020). Is there a gender gap in chemical sciences scholarly communication? *Chemical Science*, 11(8), 2277–2301. <https://doi.org/10.1039/C9SC04090K>, PubMed: 32180933
- Dehdarirad, T., & Karlsson, K. (2021). News media attention in Climate Action: Latent topics and open access. *Scientometrics*, 126(9), 8109–8128. <https://doi.org/10.1007/s11192-021-04095-7>
- Dehdarirad, T., & Sotudeh, H. (2017). The impact of information about authors and paper research topic on citation disparities: A gender perspective. *22nd Nordic Workshop on Bibliometrics and Research Policy*, Helsinki, Finland. <https://doi.org/10.6084/m9.figshare.5624758.v1>
- Didegah, F., Bowman, T. D., & Holmberg, K. (2018). On the differences between citations and altmetrics: An investigation of factors driving altmetrics versus citations for Finnish articles. *Journal of the Association for Information Science and Technology*, 69(6), 832–843. <https://doi.org/10.1002/asi.23934>
- Didegah, F., & Thelwall, M. (2013). Which factors help authors produce the highest impact research? Collaboration, journal and document properties. *Journal of Informetrics*, 7(4), 861–873. <https://doi.org/10.1016/j.joi.2013.08.006>
- Dion, M. L., & Mitchell, S. M. (2020). How many citations to women is “enough”? Estimates of gender representation in political science. *PS: Political Science & Politics*, 53(1), 107–113. <https://doi.org/10.1017/S1049096519001173>
- Dion, M. L., Sumner, J. L., & Mitchell, S. M. (2018). Gendered citation patterns across political science and social science methodology fields. *Political Analysis*, 26(3), 312–327. <https://doi.org/10.1017/pan.2018.12>
- Eysenbach, G. (2006). Citation advantage of open access articles. *PLOS Biology*, 4(5), e157. <https://doi.org/10.1371/journal.pbio.0040157>, PubMed: 16683865
- Fidrmuc, J., & Paphawasit, B. (2018). Beautiful minds: Physical attractiveness and research productivity in economics. *2nd IZA/HSE Workshop*, Moscow.
- Finch, W. H., Bolin, J. E., & Kelley, K. (2019). Longitudinal data analysis using multilevel models. In *Multilevel Modeling Using R*. CRC Press. <https://doi.org/10.1201/9781351062268-5>
- Francke, H. (2021). Trust in the academy: A conceptual framework for understanding trust on academic web profiles. *Journal of Documentation*, 78(7), 192–210. <https://doi.org/10.1108/JD-01-2021-0010>
- Gallivan, M. J. (2012). Analyzing citation impact of IS research by women and men: Do women have higher levels of research impact? *Proceedings of the 50th Annual Conference on Computers and People Research*, Milwaukee, Wisconsin. <https://doi.org/10.1145/2214091.2214137>
- Gargouri, Y., Hajjem, C., Larivière, V., Gingras, Y., Carr, L., ... Harnad, S. (2010). Self-selected or mandated, open access increases citation impact for higher quality research. *PLOS ONE*, 5(10), e13636. <https://doi.org/10.1371/journal.pone.0013636>, PubMed: 20976155
- Gelman, A., & Hill, J. (2008). *Data analysis using regression and multilevel/hierarchical models*. New York: Cambridge University Press.
- Ghiasi, G., Mongeon, P., Sugimoto, C., & Larivière, V. (2018). Gender homophily in citations. *23rd International Conference on Science and Technology Indicators*, Leiden.
- Gvozdanović, J., & Maes, K. (2018). *Implicit bias in academia: A challenge to the meritocratic principle and to women's careers – And what to do about it*. Retrieved from <https://www.leru.org/publications/implicit-bias-in-academia-a-challenge-to-the-meritocratic-principle-and-to-womens-careers-and-what-to-do-about-it>.
- Gunasekara, F. I., Richardson, K., Carter, K., & Blakely, T. (2013). Fixed effects analysis of repeated measures data. *International Journal of Epidemiology*, 43(1), 264–269. <https://doi.org/10.1093/ije/dyt221>, PubMed: 24366487
- Hajjem, C., Harnad, S., & Gingras, Y. (2006). Ten-year cross-disciplinary comparison of the growth of open access and how it increases research citation impact. *CoRR, abs/cs/0606079*. <https://doi.org/10.48550/arXiv.cs/0606079>
- Harris, M., Macinko, J., Jimenez, G., & Mullachery, P. (2017). Measuring the bias against low-income country research: An implicit association test. *Globalization and Health*, 13(1), 80. <https://doi.org/10.1186/s12992-017-0304-y>, PubMed: 29110668
- Ito, T. A., & Umland, G. R. (2003). Race and gender on the brain: Electro-cortical measures of attention to the race and gender of multiply categorizable individuals. *Journal of Personality and Social Psychology*, 85(4), 616–626. <https://doi.org/10.1037/0022-3514.85.4.616>, PubMed: 14561116
- Jarboe, L. R. (2019). Regional, institutional, and departmental factors associated with gender diversity among BS-level chemical and electrical engineering graduates. *PLOS ONE*, 14(10), e0223568. <https://doi.org/10.1371/journal.pone.0223568>, PubMed: 31596881
- Kuznetsova, A., Brockhoff, P., & Christensen, R. (2017). ImerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13), 1–26. <https://doi.org/10.18637/jss.v082.i13>
- Lüdecke, D., Ben-Shachar, M., Patil, I., Waggoner, P., & Makowski, D. (2021). performance: An R package for assessment, comparison and testing of statistical models. *Journal of Open Source Software*, 6(60), 3139. <https://doi.org/10.21105/joss.03139>
- Mahbuba, D., & Rousseau, R. (2011). The Matthew effect and a relation with concept symbols and defaults. *Annals of Library and Information Studies*, 58, 335–345.
- Merton, R. K. (1973). *The sociology of science: Theoretical and empirical investigations*. Chicago: University of Chicago Press.
- Mishra, S., Fegley, B. D., Diesner, J., & Torvik, V. I. (2018). Self-citation is the hallmark of productive authors, of any gender. *PLOS ONE*, 13(9), e0195773. <https://doi.org/10.1371/journal.pone.0195773>, PubMed: 30256792
- Mishra, S., & Wang, K. (2021). Convergence and inequality in research globalization. *arXiv:2103.02052*. <https://doi.org/10.48550/arXiv.2103.02052>
- Mutz, R., & Daniel, H.-D. (2019). How to consider fractional counting and field normalization in the statistical modeling of bibliometric data: A multilevel Poisson regression approach. *Journal of Informetrics*, 13(2), 643–657. <https://doi.org/10.1016/j.joi.2019.03.007>
- Mutz, R., Wolbring, T., & Daniel, H.-D. (2017). The effect of the “very important paper” (VIP) designation in *Angewandte Chemie International Edition* on citation impact: A propensity score matching analysis. *Journal of the Association for Information Science and Technology*, 68(9), 2139–2153. <https://doi.org/10.1002/asi.23701>
- Nielsen, M. W., & Andersen, J. P. (2021). Global citation inequality is on the rise. *Proceedings of the National Academy of Sciences*, 118(7), e2012208118. <https://doi.org/10.1073/pnas.2012208118>, PubMed: 33558230

- Nielsen, M. W., Baker, C. F., Brady, E., Petersen, M. B., & Andersen, J. P. (2021). Weak evidence of country- and institution-related status bias in the peer review of abstracts. *eLife*, *10*, e64561. <https://doi.org/10.7554/eLife.64561>, PubMed: 33734086
- Petersen, A. M., Fortunato, S., Pan, R. K., Kaski, K., Penner, O., ... Pammolli, F. (2014). Reputation and impact in academic careers. *Proceedings of the National Academy of Sciences*, *111*(43), 15316–15321. <https://doi.org/10.1073/pnas.1323111111>, PubMed: 25288774
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (Vol. 1). New Delhi: Sage.
- Ridgeway, C. L. (2011). *Framed by gender*. Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199755776.001.0001>
- Schumann, P., & Kuchinke, L. (2020). Do(n't) worry, it's temporary: The effects of fixed-term employment on affective well-being. *Journal of Happiness Studies*, *21*(7), 2557–2582. <https://doi.org/10.1007/s10902-019-00194-8>
- Schunck, R. (2013). Within and between estimates in random-effects models: Advantages and drawbacks of correlated random effects and hybrid models. *The Stata Journal*, *13*(1), 65–76. <https://doi.org/10.1177/1536867X1301300105>
- Tak, E., Correll, S. J., Soule, S. A. (2019). Gender inequality in product markets: When and how status beliefs transfer to products. *Social Forces*, *98*(2), 548–577. <https://doi.org/10.1093/sf/soy125>
- Thelwall, M. (2016). The discretised lognormal and hooked power law distributions for complete citation data: Best options for modelling and regression. *Journal of Informetrics*, *10*(2), 336–346. <https://doi.org/10.1016/j.joi.2015.12.00>
- Thelwall, M., & Sud, D. (2020). Greater female first author citation advantages do not associate with reduced or reducing gender disparities in academia. *Quantitative Science Studies*, *1*(3), 1283–1297. https://doi.org/10.1162/qss_a_00069
- Vinkler, P. (2010). Scientometric indicators for the assessment of publications. In P. Vinkler (Ed.), *The evaluation of research by scientometric indicators* (pp. 79–124). Chandos Publishing. <https://doi.org/10.1016/B978-1-84334-572-5.50007-X>
- Wang, M., Jiao, S., Zhang, J., Zhang, X., & Zhu, N. (2020). Identification high influential articles by considering the topic characteristics of articles. *IEEE Access*, *8*, 107887–107899. <https://doi.org/10.1109/ACCESS.2020.3001190>
- West, B. T., Welch, K. B., & Galecki, A. T. (2015). *Linear mixed models: A practical guide using statistical software*. Ann Arbor: CRC Press. <https://doi.org/10.1201/b17198>
- Wooldridge, J. M. (2010). *Additional single-equation topics*. In *Econometric analysis of cross section and panel data*. Boston, MA: MIT Press.