Modelling the Effects of Open Access, Gender and Collaboration on Citation Outcomes: Replicating, Expanding and Drilling

Struck, D.B.; Durning, M.; Roberge, G.; Campbell, D.

STI 2018 Conference Proceedings

Proceedings of the 23rd International Conference on Science and Technology Indicators

All papers published in this conference proceedings have been peer reviewed through a peer review process administered by the proceedings Editors. Reviews were conducted by expert referees to the professional and scientific standards expected of a conference proceedings.

Chair of the Conference

Paul Wouters

Scientific Editors

Rodrigo Costas
Thomas Franssen
Alfredo Yegros-Yegros

Layout

Andrea Reyes Elizondo
Suze van der Luijt-Jansen

The articles of this collection can be accessed at https://hdl.handle.net/1887/64521

ISBN: 978-90-9031204-0

23rd International Conference on Science and Technology Indicators (STI 2018)
"Science, Technology and Innovation indicators in transition"
12 - 14 September 2018 | Leiden, The Netherlands
#STI18LDN

Modelling the Effects of Open Access, Gender and Collaboration on Citation Outcomes: Replicating, Expanding and Drilling[1]

(David) Brooke Struck[*], Matthew Durning[**], Guillaume Roberge[**], and David Campbell[**]

[*]*brooke.struck@science-metrix.com*
Science-Metrix Inc., 1335 Mont-Royal est, Montreal, Quebec, H2J 1Y6 (Canada)

[**] *matt.durning@science-metrix.com; guillaume.roberge@science-metrix.com; david.campbell@science-metrix.com*
Science-Metrix Inc., 1335 Mont-Royal est, Montreal, Quebec, H2J 1Y6 (Canada)

ABSTRACT
Citation outcomes are often used as a proxy to measure research excellence. Accordingly, clarifying the drivers of citation outcomes is valuable information for research policy, both to inform interventions that contribute to excellence and to establish fair normalization practices when measuring the relative excellence of a body of work compared a given benchmark.

Previous research has shown that open access, international collaboration and the gender composition of research teams each have an influence on citation outcomes and that they are interconnected with one another. The present work replicates a small-scale study that disentangles their influence, expands the analysis to a much wider scope, and drills down into subsets of data, applying this analytical approach sequentially to extract valuable contextual information to situate interpretation of more localized findings.

However, this study can only inform a discussion about the policy relevance of and appropriate policy responses to the present findings; it cannot replace that discussion. Some questions that it raises pertain to the difference between a strategy to improve research and a strategy to improve citations, what counts as a fair expectation against which to measure performance for various groups, and what exactly it is that we are aiming to promote when using citation-based measures for research evaluation.

[1] This work was supported by Science-Metrix Inc.

Introduction
Previous studies have demonstrated that open access (OA), international collaboration and the gender composition of research teams are each individually connected to citation outcomes; however, these dimensions are also inter-related. Only one small-scale study to date has examined the question of whether each of these three parameters continues to have an influence on citation when controlling for the others (Struck, Roberge, & Campbell, 2017).

The present paper has three objectives.

1. Assess the robustness of the results of the previous, small-scale study by attempting to replicate the results. The replication will be conceptual rather than exact (Velden, 2017): a different data source is used here than in the original study, the indicators are parametrized in a different way, and several modelling approaches are compared.

2. Expand this analysis approach to a much wider scope. Whereas the previous study examined the dynamics of only two subfields of research, the present work assesses dynamics across all subfields (except the Arts & Humanities, see footnote 7). Furthermore, this study will cover the 2008–2012 period and all papers worldwide (rather than just papers from 2012 involving a US author).

3. Finally, "drill down" into subsamples of the overall data set to extract policy-relevant contextual information, comparing a citation benefit in a local context to the benefit "in general." This type of contextualization can move policy discussions from discussions of global trends to whether a particular intervention would be beneficial "to us" within a local context.

Methodology
Data sources
The Web of Science (WoS; produced by Clarivate Analytics) provided information on scholarly papers, including year of publication, names of authors, institutional affiliations with addresses, title of the journal in which papers appear and citation counts.

1findr (produced by 1science) is another database of scholarly literature. The contents of 1findr were matched onto the WoS to enrich the latter with data about the open access (OA) status and route of papers.

Additionally, the NamSor API was used as a data enrichment tool. NamSor uses machine learning to tag names by gender. Feeding author names from the WoS into the NamSor API provided gender tagging of author names.

Coverage and filters
The present study covers the 2008–2012 period. While papers published in 2014 or earlier could be analyzed, leaving a citation window of at least three years, embargo periods must also be considered. For example, if a paper was published behind a paywall in 2014 and available in OA only as of 2016, then there was an insufficient "OA citation window" for this paper. Accordingly, papers published after 2012 were excluded from the study.

Acknowledging that an OA citation advantage may be an evolving phenomenon, it was decided that the study would cover a five-year window, not covering papers published before 2008. This decision was further supported by potential problems related to the "backfilling effect": older papers being made available in OA well after their embargo period has elapsed.

Two filters were applied to the data. First, the lists of journals covered in the WoS and in 1findr do not overlap completely. Any journals covered in the WoS but not on the 1science white list[2] were filtered out from the data set used here. Second, papers were only considered if all authors listed could be tagged by gender using NamSor. Because about 70% of papers have two to four authors, relaxing this data quality criterion could have introduced notable noise into the analysis, as having one team member out of four improperly tagged would have dramatically shifted the recorded gender composition of the research team.[3]

Parametrizing indicators

Open access (OA) was parametrized by routes to OA: gold, green, gold + green, or unknown. Gold OA papers are those for which access is provided by the publisher, while green OA papers are those for which access is provided by another party, such as the researchers or their institution. Routes to OA were tagged based on the URL at which the papers were made available, as indexed in 1findr. Because papers can be available in multiple locations online simultaneously, some papers were available through both green and gold routes. Furthermore, as the long tail of host domains could not be tagged exhaustively, some sources remained unknown.[4]

Gender composition of research teams was binned into five categories: less than 20% women, 20%–40% women, 40%–60% women, 60%–80% women, and more than 80% women. With most papers having four authors or fewer, each permutation of gender composition fell into its own distinct bin for this number of authors; one more woman within a team of researchers always implies a change of bin with four authors or fewer and bins of this size.

The number of authors and the number of institutional addresses were retained as scalar variables. Following the findings of the previous study, international collaboration was parametrized as a binary variable rather than as a scalar number of countries. To account for differences in citation practice over time and across areas of research, citation scores were normalized by year and subfield[5] of publication.

Modelling approaches

Three regression modelling approaches were applied in this study: negative binomial, robust and zero-inflated negative binomial. The negative binomial model treats the outcome variable—citations—as count data; this treatment was warranted because citation counts are always positive whole numbers. Furthermore, as the detected variance was much larger than the mean, regression models based on the negative binomial distribution form were preferable to the Poisson distribution.

Robust regression is resilient against input and output variables that are non-normally distributed, another notable trait of these data. Because the majority of citations accrue to a small minority of papers, individual scores were log-normalized (for the robust regression) in order to mitigate problems of non-normal distribution.

[2] No public documentation is presently available on this white list. However, only a small number of journals indexed in the WoS are not also on the 1science list; as shown in Table 8, the effect of the 1science filter is small in magnitude.

[3] Unlike the previous study, gender tags were considered even if NamSor could not provide a 100% confidence ranking to the tag. The effects of these two filters on the data set composition are shown in Table 8.

[4] These were most likely green OA sources, because they are more likely than publisher websites to account for only a small volume of papers and thus find themselves in the long tail.

[5] Using the Science-Metrix taxonomy of science: http://science-metrix.com/classification

Finally, the zero-inflated negative binomial regression is useful in situations where there are many instances of zero counts, and there were many uncited papers in the data set. The zero-inflated model made it possible to estimate zero citations from two sources: the count model and a binomial model for estimating the excess zeroes.[6]

Analysis

Objective 1: Replicating previous findings in two subfields

The previous study focused on two subfields: Cardiovascular System & Hematology, and Developmental Biology. That analysis drew the following conclusions:

a) OA had a significant and meaningful effect on citation scores, in both subfields.

b) International collaboration had a significant and a meaningful effect (albeit smaller than that for OA) in Cardiovascular System & Hematology. In Developmental Biology the effect was not statistically significant.

c) Gender composition had no significant effect in Cardiovascular System & Hematology. In Developmental Biology, the effect was statistically significant, with citation scores highest for mixed-gender teams. Because of the way in which the indicator was parametrized, identifying the optimal gender balance (optimal from a citation perspective) was not possible in a thorough fashion, though mixed-leaning-male seemed to have the highest citation effect. The effect size was much smaller for gender composition than it was for OA.

The replication presented here reached broadly similar conclusions:

a) OA had large and statistically significant effects in both subfields. In Cardiovascular System & Hematology, the effect of gold OA was particularly pronounced. These findings held across all three models.

b) International collaboration had a significant effect in both subfields, though it was smaller in Developmental Biology, where previous results were not statistically significant. Furthermore, the effect of international collaboration was once again smaller than that of OA. These findings held across all three models.

c) Gender composition showed statistically significant results across all three models in Developmental Biology: the more women involved in the research team, the lower the citation score was predicted to be, with decreases of up to 30%. This finding contradicted previous results, which suggested (based on a very coarse approach to parametrization) that mixed-gender teams were optimal. In Cardiovascular System & Hematology, only the robust regression model showed statistically significant results: an increase in citations with an increase in the proportion of women in the research team.

[6] The results of the logistic regression model are not interpreted here.

Table 1: Determinants of citation score in Developmental Biology

	Variable	Type	Predicted benefit			Coefficients			Stat. significance		
			Negative binomial	Robust regression	Zero-inflated	Negative binomial	Robust regression	Zero-inflated	Negative binomial	Robust regression	Zero-inflated
Open access	Gold	Binary	38%	66%	38%	0.323	0.505	0.321	p < 2.00e-16	p < 2.00e-16	p < 2.00e-16
	Green	Binary	48%	45%	47%	0.390	0.374	0.388	p < 2.00e-16	p < 2.00e-16	p < 2.00e-16
	Green + Gold	Binary	60%	71%	59%	0.497	0.539	0.465	p < 2.00e-16	p < 2.00e-16	p < 2.00e-16
	Unknown	Binary	46%	35%	46%	0.378	0.302	0.377	p < 2.00e-16	p < 2.00e-16	p < 2.00e-16
% women in team	<20%	Binary	0%	0%	0%	n/a	n/a	n/a	n/a	n/a	n/a
	20%–40%	Binary	-13%	-7%	-13%	-0.145	-0.070	-0.145	p < 2.00e-16	p = 1.61e-07	p < 2.00e-16
	40%–60%	Binary	-20%	-12%	-20%	-0.218	-0.128	-0.218	p < 2.00e-16	p < 2.00e-16	p < 2.00e-16
	60%–80%	Binary	-33%	-20%	-33%	-0.400	-0.226	-0.400	p < 2.00e-16	p < 2.00e-16	p < 2.00e-16
	>80%	Binary	-31%	-22%	-31%	-0.377	-0.254	-0.372	p < 2.00e-16	p < 2.00e-16	p < 2.00e-16
Collab.	Number of authors	Scalar	4%	6%	4%	0.042	0.054	0.042	p < 2.00e-16	p < 2.00e-16	p < 2.00e-16
	Number of addresses	Scalar									
	International collab.	Binary	15%	16%	15%	0.140	0.151	0.140	p < 2.00e-16	p < 2.00e-16	p < 2.00e-16

Table 2: Determinants of citation score in Cardiovascular System & Hematology

	Variable	Type	Predicted benefit			Coefficients			Stat. significance		
			Negative binomial	Robust regression	Zero-inflated	Negative binomial	Robust regression	Zero-inflated	Negative binomial	Robust regression	Zero-inflated
Open access	Gold	Binary	113%	124%	111%	0.757	0.805	0.748	p < 2.00e-16	p < 2.00e-16	p < 2.00e-16
	Green	Binary	57%	74%	55%	0.454	0.556	0.438	p < 2.00e-16	p < 2.00e-16	p < 2.00e-16
	Green + Gold	Binary	145%	147%	144%	0.896	0.906	0.894	p < 2.00e-16	p < 2.00e-16	p < 2.00e-16
	Unknown	Binary	44%	50%	43%	0.366	0.404	0.355	p < 2.00e-16	p < 2.00e-16	p < 2.00e-16
% women in team	<20%	Binary	0%	0%	0%	n/a	n/a	n/a	n/a	n/a	n/a
	20%–40%	Binary	2%	9%	2%	0.021	0.085	0.017	p = 4.59e-02	p = 2.26e-12	p = 1.08e-01
	40%–60%	Binary	3%	13%	2%	0.033	0.126	0.024	p = 8.77e-03	p < 2.00e-16	p = 6.16e-02
	60%–80%	Binary	1%	14%	0%	0.013	0.131	-0.001	p = 4.92e-01	p = 1.85e-09	p = 9.41e-01
	>80%	Binary	4%	18%	2%	0.040	0.163	0.021	p = 1.64e-01	p = 1.77e-06	p = 4.64e-01
Collab.	Number of authors	Scalar	4%	6%	4%	0.038	0.060	0.036	p < 2.00e-16	p < 2.00e-16	p < 2.00e-16
	Number of addresses	Scalar	5%	5%	5%	0.053	0.051	0.051	p < 2.00e-16	p < 2.00e-16	p < 2.00e-16
	International collab.	Binary	27%	39%	26%	0.242	0.330	0.235	p < 2.00e-16	p < 2.00e-16	p < 2.00e-16

Objective 2: Expanding the scope of analysis

All research worldwide

Looking beyond these two subfields, what did we find in our analysis across all areas of research[7] over the 2008–2012 period? Once again, OA was the strongest predictor of citation outcomes. While gold OA offered only an increase of 10%–15% in citation score, the benefit of green OA was 55%–65% across our three modelling approaches. Papers available in both green and gold OA enjoyed a benefit somewhat lower than green OA alone, at about 45%. Unknown OA was similar to green + gold OA (noting that unknown OA sources were most likely to be green OA).

Turning to gender composition of research teams, the negative binomial and zero-inflated models suggested that citation scores decreased as the share of women involved increased. The robust regression model predicted a miniscule citation benefit for mixed-gender teams (again highest for male-dominated teams, as the previous study suggested), though the coefficients and statistical significance of these findings, and their inconsistency with the other two models, render this finding highly dubious. The effect sizes of gender composition were much smaller than those for OA: the largest variance was a 12% decrease in citations for all-female teams relative to all-male teams.

Looking finally at collaboration, the three models presented consistent findings. The number of researchers and the number of institutions listed in authorship were associated with citation score increases of a few percent each. International collaboration offered a benefit of 17%–23%.

Objective 3: "Drilling down" into local contexts

The present section explores the determinants of citation score in subsets of the data set modelled in Table 5.[8]

Environmental Science worldwide

The three models were unanimous: OA offered a smaller citation advantage in Environmental Science than elsewhere. In fact, gold OA came with a citation penalty, while the advantage of green OA was smaller than elsewhere.

The picture for gender composition was clear too: the greater the share of women involved, the lower the predicted citation score. This result was consistent across models, and broadly in line with the findings across all thematic areas. International collaboration offered a citation benefit of about 10%–20%, consistently smaller than the benefit in other research topics.

[7] Arts & Humanities were excluded from the analysis because single authorship is such a prevalent phenomenon in these areas. Because gender composition and international collaboration were key areas of focus in this study, and areas parametrized through co-authorship, the present approach is not a comfortable fit with single-author papers.

[8] The sample size of these subsets is presented in Table 8.

Table 3: Determinants of citation score for research worldwide across all fields (excl. Arts and Humanities)

	Variable	Type	Predicted benefit			Coefficients			Stat. significance		
			Negative binomial	Robust regression	Zero-inflated	Negative binomial	Robust regression	Zero-inflated	Negative binomial	Robust regression	Zero-inflated
Open access	Gold	Binary	15%	11%	15%	0.142	0.106	0.142	p < 2.00e-16	p < 2.00e-16	p < 2.00e-16
	Green	Binary	55%	65%	54%	0.436	0.499	0.429	p < 2.00e-16	p < 2.00e-16	p < 2.00e-16
	Green + Gold	Binary	44%	45%	44%	0.367	0.372	0.362	p < 2.00e-16	p < 2.00e-16	p < 2.00e-16
	Unknown	Binary	45%	47%	45%	0.374	0.385	0.369	p < 2.00e-16	p < 2.00e-16	p < 2.00e-16
% women in team	<20%	Binary	0%	0%	0%	n/a	n/a	n/a	n/a	n/a	n/a
	20%–40%	Binary	-3%	3%	-4%	-0.031	0.033	-0.036	p < 2.00e-16	p < 2.00e-16	p < 2.00e-16
	40%–60%	Binary	-6%	1%	-6%	-0.061	0.006	-0.064	p < 2.00e-16	p = 6.60e-03	p < 2.00e-16
	60%–80%	Binary	-10%	-1%	-11%	-0.107	-0.010	-0.113	p < 2.00e-16	p = 1.68e-04	p < 2.00e-16
	>80%	Binary	-12%	-6%	-12%	-0.128	-0.057	-0.123	p < 2.00e-16	p < 2.00e-16	p < 2.00e-16
Collab.	Number of authors	Scalar	3%	4%	2%	0.025	0.041	0.023	p < 2.00e-16	p < 2.00e-16	p < 2.00e-16
	Number of addresses	Scalar	3%	3%	4%	0.034	0.025	0.035	p < 2.00e-16	p < 2.00e-16	p < 2.00e-16
	International collab.	Binary	18%	23%	17%	0.162	0.208	0.161	p < 2.00e-16	p < 2.00e-16	p < 2.00e-16

Table 4: Determinants of citation score for research worldwide in Environmental Science

	Variable	Type	Predicted benefit			Coefficients			Stat. significance		
			Negative binomial	Robust regression	Zero-inflated	Negative binomial	Robust regression	Zero-inflated	Negative binomial	Robust regression	Zero-inflated
Open access	Gold	Binary	-12%	-30%	-11%	-0.128	-0.354	-0.119	p = 3.21e-16	p < 2.00e-16	p = 2.74e-13
	Green	Binary	36%	38%	36%	0.308	0.319	0.306	p < 2.00e-16	p < 2.00e-16	p < 2.00e-16
	Green + Gold	Binary	21%	4%	21%	0.190	0.043	0.188	p < 2.00e-16	p = 4.45e-02	p < 2.00e-16
	Unknown	Binary	44%	41%	44%	0.366	0.347	0.364	p < 2.00e-16	p < 2.00e-16	p < 2.00e-16
% women in team	<20%	Binary	0%	0%	0%	n/a	n/a	n/a	n/a	n/a	n/a
	20%–40%	Binary	-3%	0%	-3%	-0.032	0.000	-0.033	p = 5.36e-05	p = 9.73e-01	p = 2.81e-05
	40%–60%	Binary	-5%	-3%	-5%	-0.050	-0.030	-0.049	p = 2.34e-08	p = 2.26e-03	p = 5.64e-08
	60%–80%	Binary	-8%	-5%	-8%	-0.084	-0.055	-0.085	p = 6.11e-10	p = 1.70e-04	p = 5.84e-10
	>80%	Binary	-13%	-11%	-13%	-0.142	-0.122	-0.139	p = 9.25e-16	p = 2.00e-09	p = 7.77e-15
Collab.	Number of authors	Scalar	5%	6%	4%	0.045	0.054	0.044	p < 2.00e-16	p < 2.00e-16	p < 2.00e-16
	Number of addresses	Scalar	2%	2%	2%	0.024	0.019	0.024	p = 6.48e-16	p = 7.2.e-09	p < 2.00e-16
	International collab.	Binary	11%	18%	10%	0.100	0.165	0.098	p < 2.00e-16	p < 2.00e-16	p < 2.00e-16

Canadian research
The picture once again changed when looking at Canadian research. There was clearly an OA citation advantage, though the gap between gold OA and green OA was about 15–20 percentage points, much smaller here than for all research worldwide. The citation advantage of gold OA was nearly 25%, about 2–5 percentage points higher for Canadian research than all research globally. The advantage of green OA was also about 15–20 percentage points lower for Canada than worldwide.

For Canadian research, all-female research teams were predicted to have citation scores about 4%–8% lower than all-male research teams. This gap was smaller than for research worldwide. However, mixed-gender teams fared worse in the Canadian context than they did generally across the globe, by 1–4 percentage points. Estimates for gender composition results were statistically significant in most but not all cases in this subset.

In terms of international collaboration, Canadian researchers enjoyed a 15% citation advantage when partnering with co-authors abroad, a level similar to the worldwide pattern, except according to the robust regression.

Canadian Environmental Science
Gold OA came with a substantial citation penalty in this subset: papers received one third fewer citations than papers behind paywalls, an even larger penalty than for gold OA in Environmental Science worldwide. Furthermore, Canadian research (across thematic areas) generally enjoyed a relatively large gold OA citation advantage; the gold OA penalty for Canadian Environmental Science was considerable.

Green OA offered a citation advantage of about 25% for Canadian Environmental Science. This advantage was about 15 percentage points smaller than for Environmental Science worldwide, about 15 percentage points smaller than for all Canadian research across all thematic areas, and about 30 percentage points smaller than the green OA citation advantage for all research globally.

Looking at the effects of gender composition, the statistical significance of the findings was severely degraded in this subset of the data. The sign and magnitude of the estimates were consistent across the three models, though: female-dominated teams seemed to fare better than balanced or male-dominated teams. However, there were so few observations in these bins (see Table 9) that the results are unlikely to be reliable. While the consistency across models may pique curiosity, further replication and theoretical explanation of these phenomena would be needed before much stock should be put in these findings about gender composition in Canadian Environmental Science.

Finally, international collaboration in Canadian Environmental Science offered a citation benefit of more than 20%. The benefit here was about 5 percentage points higher than for Canadian research across thematic areas, about 10 percentage points higher than in Environmental Science worldwide, and about 5 percentage points higher than for all research globally. Canadian Environmental Science was the only area studied here in which international collaboration and green OA offered benefits of similar magnitude; in all other instances, green OA offered a much larger benefit.

Table 5: Determinants of citation score for Canadian research across all fields (excl. Arts and Humanities)

	Variable	Type	Predicted benefit			Coefficients			Stat. significance		
			Negative binomial	Robust regression	Zero-inflated	Negative binomial	Robust regression	Zero-inflated	Negative binomial	Robust regression	Zero-inflated
Open access	Gold	Binary	15%	12%	15%	0.139	0.117	0.140	p < 2.00e-16	p < 2.00e-16	p < 2.00e-16
	Green	Binary	40%	40%	40%	0.337	0.338	0.334	p < 2.00e-16	p < 2.00e-16	p < 2.00e-16
	Green + Gold	Binary	31%	28%	30%	0.267	0.247	0.265	p < 2.00e-16	p < 2.00e-16	p < 2.00e-16
	Unknown	Binary	42%	40%	42%	0.353	0.333	0.351	p < 2.00e-16	p < 2.00e-16	p < 2.00e-16
% women in team	<20%	Binary	0%	0%	0%	n/a	n/a	n/a	n/a	n/a	n/a
	20%–40%	Binary	-5%	0%	-5%	-0.051	-0.003	-0.053	p = 2.58e-11	p = 7.26e-01	p = 3.86e-12
	40%–60%	Binary	-8%	-3%	-8%	-0.079	-0.034	-0.081	p < 2.00e-16	p = 8.07e-05	p < 2.00e-16
	60%–80%	Binary	-11%	-3%	-11%	-0.115	-0.031	-0.118	p < 2.00e-16	p = 4.91e-03	p < 2.00e-16
	>80%	Binary	-9%	-2%	-8%	-0.089	-0.022	-0.088	p = 9.29e-14	p = 9.24e-02	p = 1.50e-13
Collab.	Number of authors	Scalar	4%	5%	4%	0.040	0.050	0.040	p < 2.00e-16	p < 2.00e-16	p < 2.00e-16
	Number of addresses	Scalar	4%	2%	4%	0.035	0.021	0.036	p < 2.00e-16	p < 2.00e-16	p < 2.00e-16
	International collab.	Binary	17%	16%	17%	0.155	0.149	0.155	p < 2.00e-16	p < 2.00e-16	p < 2.00e-16

Table 6: Determinants of citation score for Canadian research in Environmental Science

	Variable	Type	Predicted benefit			Coefficients			Stat. significance		
			Negative binomial	Robust regression	Zero-inflated	Negative binomial	Robust regression	Zero-inflated	Negative binomial	Robust regression	Zero-inflated
Open access	Gold	Binary	-33%	-40%	-34%	-0.406	-0.510	-0.410	p = 1.21e-06	p = 5.94e-06	p = 7.71e-07
	Green	Binary	23%	22%	22%	0.205	0.200	0.201	p = 3.38e-13	p = 1.02e-11	p = 8.46e-13
	Green + Gold	Binary	11%	22%	16%	0.106	0.199	0.145	p = 1.90e-1	p = 1.95e-02	p = 8.07e-02
	Unknown	Binary	24%	17%	24%	0.216	0.161	0.212	p = 2.84e-08	p = 2.01e-04	p = 4.70e-08
% women in team	<20%	Binary	0%	0%	0%	n/a	n/a	n/a	n/a	n/a	n/a
	20%–40%	Binary	-9%	-5%	-10%	-0.098	-0.051	-0.102	p = 1.47e-03	p = 1.11e-01	p = 9.20e-04
	40%–60%	Binary	-9%	-5%	-10%	-0.096	-0.051	-0.100	p = 7.47e-03	p = 1.83e-01	p = 4.74e-03
	60%–80%	Binary	10%	12%	9%	0.091	0.116	0.086	p = 9.61e-02	p = 4.48e-02	p = 1.15e-01
	>80%	Binary	4%	2%	4%	0.043	0.024	0.036	p = 6.12e-01	p = 8.10e-01	p = 6.68e-01
Collab.	Number of authors	Scalar	5%	7%	5%	0.052	0.067	0.050	p = 1.58e-08	p = 5.25e-12	p = 6.70e-08
	Number of addresses	Scalar	4%	2%	4%	0.036	0.019	0.037	p = 4.56e-03	p = 1.54e-01	p = 3.61e-03
	International collab.	Binary	22%	23%	22%	0.197	0.211	0.200	p = 1.15e-11	p = 1.67e-11	p = 5.27e-12

Discussion
Using a different data source, different parametrization of indicators and several modelling approaches, most previous findings were supported for Developmental Biology and Cardiovascular System & Hematology. The influence of open access and gender composition of research teams were comparable across studies, in both the signs and relative magnitudes of their influence. International collaboration had notable positive effects in both subfields, though the previous study detected none for Developmental Biology.

The wider analysis showed that the OA citation advantage holds generally, with green OA offering a much larger benefit than gold OA. The larger the proportion of women in research is, the lower the predicted citation outcomes. International collaboration offers a citation advantage, smaller than the advantage of either green or gold OA.

Drilling down into subsets of data demonstrated the potential of this approach to provide context for interpretation. For instance, while international collaboration and green OA offer a similar citation advantage in Canadian Environmental Science, this advantage represents a much stronger effect for international collaboration and a much weaker effect for green OA than are observed generally.

Conclusion
Notable policy implications of this research remain to be explored. The measurement of these various citation advantages can be used to establish more flexible benchmarks for comparing different bodies of work. As noted in the previous study, these discussions raise questions about what might be considered a strategy to increase research excellence as opposed to a strategy to "game" the assessment system.

Citations are often used as a proxy for excellence, novelty, influence or research that is ground-breaking, transformative, leading-edge or at the frontier of knowledge—though these terms are seldom given a precise definition. We know that citations also track other properties; do we value those properties too, or do we wish to filter them out as noise?

Should citation levels be relativized for OA status if we are seeking research excellence? Should we hold research to a different standard depending on the gender composition of the research team, or international collaboration? What perverse incentives (for various actors) might be introduced into the system if we introduced flexible benchmarks?

Depending on how we address these questions, we should consider the relevance of the present findings for various decisions around hiring, granting, tenure & promotion and other assessments of research and researchers. Appropriate policy responses should be developed, implemented, tested and refined. These questions are raised by the present study, not resolved by it. Indicator research contributes to this important conversation, it doesn't conclude it.

References
Struck, D. B., Roberge, G., & Campbell, D. (2017). The influence of open access, gender and co-authorship on citation scores. Presented at the Science, Technology & Innovation (STI) Indicators 2017, Paris, France. Retrieved from https://sti2017.paris/wp-content/uploads/2017/11/oa-db-struck-et-al.pdf

Velden, T. (2017, November 13). Reproducible scientometrics research: open data, code, and education. Retrieved from http://www.issi-

society.org/blog/posts/2017/november/reproducible-scientometrics-research-open-data-code-and-education-issi-2017/

Appendix tables

Table 7: Inter-correlation of model variables across all fields (excl. Arts and Humanities)

Model variables	OA type	Team gender composition	Number of authors	Number of institutions	International collaboration
OA type	100%	0%	0%	1%	1%
Team gender composition	0%	100%	0%	0%	1%
Number of authors	0%	0%	100%	38%	3%
Number of institutions	1%	0%	38%	100%	12%
International collaboration	1%	1%	3%	12%	100%

Table 8: Effects of filters on sample data set for modelling exercise

Geographic coverage	Thematic coverage	Unfiltered data set (n=)	Papers removed		Final data set (n=)	Uncited papers (% of total)
			1science white list	Gender tagging		
World	All areas	6,073,672	22,418	3,084,445	2,987,526	233,646 (7.8%)
World	Enviro sci	214,675	182	115,875	98,800	4,001 (4.0%)
World	Cardi Sys & Hema	111,042	417	26,952	47,048	5,783 (12.3%)
World	Dev Biology	74,000	626	41,820	69,222	587 (0.8%)
Canada	All areas	276,805	672	115,916	131,088	6,398 (4.9%)
Canada	Enviro sci	13,001	3	5,984	5,595	138 (2.5%)

Table 9: Number of uncited papers in each gender bin, for Tables 1–6

	Gender composition bin	Number of uncited papers (% of total)					
		Table 1	Table 2	Table 3	Table 4	Table 5	Table 6
% women in team	<20%	165 (1.1%)	3,189 (9.4%)	112,844 (8.8%)	2030 (4.2%)	3,208 (5.9%)	83 (2.9%)
	20%–40%	118 (0.9%)	1,358 (7.2%)	42,289 (6.1%)	775 (3.2%)	1,016 (3.4%)	23 (1.6%)
	40%–60%	139 (1.2%)	804 (7.4%)	41,494 (7.2%)	726 (4.4%)	1,165 (4.5%)	21 (2.3%)
	60%–80%	91 (1.8%)	297 (7.3%)	17,885 (6.7%)	255 (4.2%)	433 (3.6%)	6 (1.9%0
	>80%	74 (2.9%)	135 (8.1%)	19,134 (10.9%)	215 (6.2%)	576 (6.3%)	5 (3.9%)