AGU100 ADVANCING EARTH AND SPACE SCIENCE

JGR

Journal of Geophysical Research: Space Physics

COMMENTARY

10.1002/2018JA025291

Key Points:

- Large collaborative and international teams that cite the literature extensively write high-citation papers
- No gender bias is found in terms of citation rates between female and male first-author papers, and they submit first-author papers proportionally to their representation in the discipline
- A statistically significant small difference in citations is found for papers from U.S. institutions compared to the rest of the world

Correspondence to:
M. B. Moldwin,
mmoldwin@umich.edu;
liemohn@umich.edu

Citation:
Moldwin, M. B., & Liemohn, M. W. (2018). High-citation papers in space physics: examination of gender, country, and paper characteristics. *Journal of Geophysical Research: Space Physics*, *123*, 2557–2565. https://doi.org/10.1002/2018JA025291

Received 31 JAN 2018
Accepted 19 MAR 2018
Accepted article online 23 MAR 2018
Published online 19 APR 2018
Corrected 9 MAY 2018

This article was corrected on 9 MAY 2018. See the end of the full text for details.

High-Citation Papers in Space Physics: Examination of Gender, Country, and Paper Characteristics

Mark B. Moldwin[1] and **Michael W. Liemohn**[1]

[1]Climate and Space Sciences and Engineering, University of Michigan, Ann Arbor, MI, USA

Abstract The number of citations to a refereed journal article from other refereed journal articles is a measure of its impact. Papers, individuals, journals, departments, and institutions are increasingly judged by the impact they have in their disciplines, and citation counts are now a relatively easy (though not necessarily accurate or straightforward) way of attempting to quantify impact. This study examines papers published in the Journal of Geophysical Research—Space Physics in the year 2012 (n = 705) and analyzes the characteristics of high-citation papers compared to low-citation papers. We find that high-citation papers generally have a large number of authors (>5) and cite significantly more articles in the reference section than low-citation papers. We also examined the gender and country of institution of the first author and found that there is not a statistically significant gender bias, but there are some significant differences in citation statistics between articles based on the country of first-author institution.

Plain Language Summary The number of citations to a refereed journal article from other refereed journal articles is a measure of its impact. Papers, individuals, journals, departments, and institutions are increasingly judged by the impact they have in their disciplines, and citation counts are now a relatively easy (though not necessarily accurate) way of attempting to quantify impact. This study examines papers published in the Journal of Geophysical Research—Space Physics and analyzes the characteristics of high-citation papers compared to low-citation papers. We find that high-citation papers generally have large number of authors (>5) and cite significantly more articles in the reference section than low-citation papers. We also found that there is not a statistically significant gender bias in terms of citation counts, but there are some significant differences in citation statistics between articles based on the country of first-author institution.

1. Introduction

Bibliometrics—the use of citation data—to evaluate the impact of papers, individuals, journals, institutions, and countries is now standard practice across most disciplines (e.g., Bornmann & Daniel, 2007, 2008). Interest and usage of bibliometric values has taken off with the development of the h-index (Hirsch, 2005) and Journal Impact Factor (e.g., Garfield, 2006), though there has been pushback on the primacy of bibliometric data for assessment (e.g., San Francisco DORA, 2012).

A number of studies have examined "highly cited" papers in a variety of disciplines for the purpose of developing a "review" paper bibliography of a field's key works (e.g., Tang et al., 2016). Several different research fields have also examined characteristics of highly cited papers and found significant differences between disciplines (e.g., Puuska et al., 2014; Waltman & van Eck, 2013). A recent study examined "content" differences between highly cited and low-citation papers and found that highly cited papers contain discussions of both old and new research (Mukherjee & Romero, 2017). They found the "hot spot" for high-citation papers was when a paper's cited references (CRs) had a low mean age (e.g., preferentially cited "new" work) and a high mean age variance (also cited seminal work from the past). There has also been a number of studies that examined characteristics of highly cited papers based on a number of parameters that seemingly have no connection with the "quality or significance" of the work such as the use of different punctuation in the title (such as colons, question marks, or hyphens), the number of authors, or the commonality between title words and key words (e.g., Haustein et al., 2015; Jacques & Sebire, 2010). For colons in the title, it has been hypothesized that they may indicate scholarly complexity and distinction (Haslam et al., 2008), though it could be that compound titles allow more keywords in the title, enabling better discoverability using title searches. Puuska et al. (2014) found that more authors and, in particular, more international authors, increased the citation of the papers in their study, which spanned a number of fields including natural sciences and engineering. Clearly, there is great interest in discovering the features common to highly cited papers.

There have also been bibliometric analyses that assess the bias of a research community. Specifically, there have been a number of studies examining gender differences in publishing and citations in a variety of fields and for the most part found that men publish and are cited more than their female counterparts (e.g., Caplar et al., 2017; Van den Besselaar & Sandström, 2016). Examining astronomy publications, a field not far from space physics, Caplar et al. (2017) found that a female first-authored paper receives 10.4 ± 0.9% fewer citations over its lifetime than the expected value if the paper had been first-authored by a man (controlling for nongender factors). There are some fields, though (women's studies and library and information science), that do not show this gender difference (e.g., Penas & Willett, 2006). As research and scientific publishing becomes more international, several studies have also examined differences in output and impact based on country of the first author's institution as well as differences between papers with authors from single countries compared to international teams (e.g., Sin, 2011). This and other studies found that international coauthorship is related to higher citations. A possible explanation is that international teams submit papers to higher impact international journals, thus increasing the visibility of the work. Such studies can help identify any implicit or explicit bias and could serve as discussion starters for changing research community culture or even scientific society policy.

Because of the differences between research communities in citation practice, it is useful to examine the bibliometric trends in our field. This paper examines citation statistics for the Space Physics community, examining papers from a single year (2012) and from a single journal (*JGR Space Physics*) to determine the characteristics of high-citation papers particularly in comparison to low-citation papers. The goal is to examine how the sociology of science may impact citations, but we also include results from paper characteristics (such as title length). Where we make inferences on mechanisms, extreme caution should be used since, clearly, correlation does not indicate causation and true predictive bibliometric studies are not only difficult to conduct but also their validity are suspect (e.g., Acuna et al., 2012). However, determining if there are any differences found based on size of author list, gender, or country of institution of the first author enables the community to honestly assess its own implicit and explicit bias. This data may help editors, authors, referees, and hiring and promotion committee members to more effectively use bibliometric data and reduce or eliminate differences not based on the quality of the work.

2. Methodology

We analyze the papers published in the American Geophysical Union (AGU) Journal of Geophysical Research—Space Physics (JGR) during the year 2012 to understand characteristics of highly cited papers compared to papers with few citations. We conducted a Clarivate Analytics (formerly Thomsen-Reuters Institute for Scientific Information) Web of Science (WoS) citation count on 3 June 2016 for the cumulative citations to papers published in JGR in 2012. The citation count for each paper is the number of citations accumulated from publication in 2012 to 3 June 2016. We gathered data on a variety of parameters including number of authors and number of CRs contained in each article.

We also gathered characteristics of titles that included (1) length of title, (2) if it contained acronyms or not, (3) if it used a colon to create a subtitle, and (4) if it contained a geophysical name (such as ionosphere) and if it was used as a noun or adjective (ionospheric).

We also compiled demographic data of the authors (gender of first author, country of first author's institution, and the total number of institutions and countries represented by all the authors). For gender data, we conducted online searches of the first authors to determine gender through institutional Web pages, news articles describing their work that used gender pronouns, and our personal knowledge of many of the authors. We were able to identify the gender for 653 of the 705 first authors (note that several authors published multiple first-author papers in JGR in 2012 so each record is not necessarily a unique individual).

3. Results

Figure 1 presents a histogram of citations for the 705 papers considered in this study. The bin width is one, with the first bin shown being those papers with zero citations. Because the publication dates range throughout 2012 and the citation data were generated on 3 June 2016, the "age" of the papers ranges from 3.5 to 4.5 years, with a relatively constant number published per month, so the average age of the papers in the study is just under 4 years. Several statistics about the citation data are listed on the figure. While the

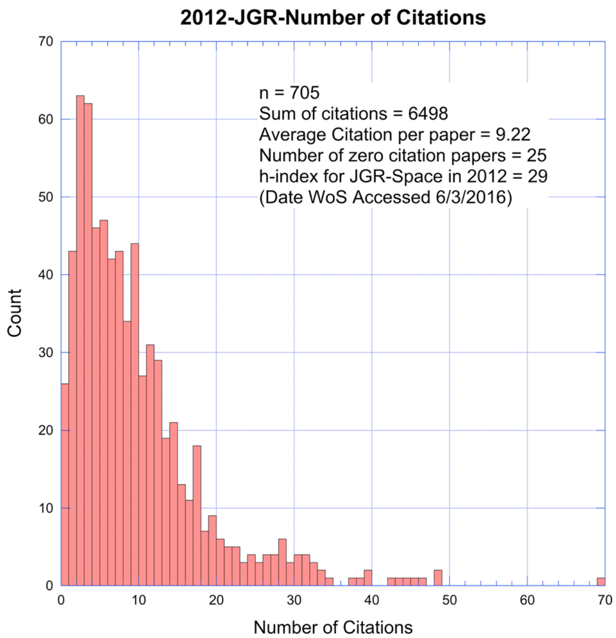


Figure 1. The distribution of citations for the 705 papers that appeared in Journal of Geophysical Research—Space Physics in 2012 from Thomsen-Reuters Institute for Scientific Information Web of Science accessed on 3 June 2016. The median citation count is 7, while the mean is 9.22.

average number of citations per paper is 9.22, the median is 7 and the mode is 2. That is, the distribution is not Gaussian, but rather skewed, with a hard cutoff at zero and a long positive tail. Also listed on Figure 1 is the h-index for JGR Space Physics for articles published in 2012, as calculated from the citations through 3 June 2016. Of the 705, 29 papers have received 29 or more citations.

It is worth noting that, after only ~4 years since publication, only 3.5% of papers have zero citations. This is much lower than the typical journal in the WoS database. Garfield (2006) presented the distribution of citations for the 38.1 million journal articles in the WoS database (spanning across a wide range of disciplines) that were published in 1990 through 2005. They found that 19.9 million articles received 1 or more citations as of late 2005, indicating that 48% of papers in this database had zero citations.

3.1. Cited References

Earlier studies have found that highly cited papers cite more papers than papers that are less cited (Fox et al., 2016). Figure 2 (left) examines the correlation of citations versus the number of CRs, and Figure 2 (middle) shows the same data arranged in deciles of citations (i.e., each bin along the *x* axis has approximately 10% of the distribution sorted from low CR to high CR). The vertical and horizontal error bars show the standard deviation within the bin, and the lines are the best fit to the data and means in each panel, respectively. The linear best fit line shows a significant correlation between increasing citation count with increasing number of CRs; however, the figures clearly show that there is significant scatter and that outliers play significant role in the trends.

Another way to look at the significance of the correlation of citations to CRs is to look at the two distributions of papers that have above and below the median number of CRs. For the papers in our study, the median CR number is 39. The right panel of Figure 2 shows the whisker plot of the distribution of citations for these two populations. The first two lines of data in Table 1 shows the statistics of the two distributions and results from a Welch's student *t* test, which is appropriate for unpaired data with unequal variance. The distributions are clearly different ($P(t) < 0.0001$), showing that papers above the median in terms of CRs receive more citations than papers that have fewer CRs.

The results from Figure 2 and Table 1 demonstrate that there is a statistically significant difference between citation counts of papers with high numbers of CR compared to small numbers of CR. There is significant

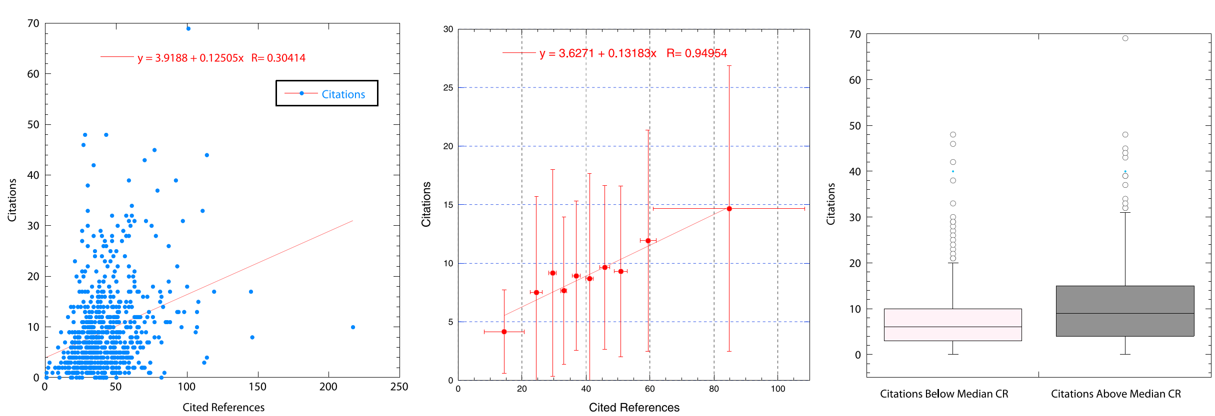


Figure 2. (left) The relationship between citations and cited references for the distribution of papers with a linear fit to the data. (middle) The relationship of the mean cited references versus citation count. The error bars, both vertical and horizontal, represent the standard deviation about the mean in each bin. (right) The whisker plot of the citation statistics for the two populations (above and below the median citation count). The box line shows the median citation, and the box shows the width of the quartile ranges.

Table 1
Parameter Subset Statistics of Times Cited and Welch's T Test Comparisons Between Subsets

	Count	Mean	Variance	St. dev.	St. err.	Mean diff	Deg. of freedom	*T* value	*P*(*t*)
Below median # of cited refs	355	7.428	49.89	7.063	0.3749	−3.592	645	−5.708	<0.0001
Above median # of cited refs	350	11.02	89.43	9.457	0.5055				
Below median # of authors	357	7.992	59.08	7.686	0.4068	−2.471	676	−3.878	0.000115
Above median # of authors	348	10.463	83.68	9.148	0.4904				
Below median # of institutions	429	8.576	63.90	7.994	0.3859	−1.872	507	−2.742	0.006315
Above median # of institutions	268	10.448	84.96	9.217	0.5630				
Below median # of countries	553	8.814	64.48	8.030	0.3415	−2.332	193	−2.582	0.01056
Above median # of countries	144	11.146	100.69	10.034	0.8362				
USA first author	321	10.184	77.69	8.814	0.4920	1.626	660	2.504	0.01252
Rest of world first author	375	8.557	67.44	8.212	0.4241				
Male first author	528	9.462	76.11	8.724	76.112	0.301	184	−0.344	0.731
Female first author	124	9.161	76.79	8.726	76.787				
Below median # title words	370	9.462	83.18	9.120	0.4741	0.5278	700	0.8272	0.4084
Above median # title words	335	8.934	61.09	7.816	0.4270				
No acronyms in the title	463	9.229	72.80	8.532	0.397	0.0513	489	0.0758	0.940
One or more acronym in title	242	9.177	72.66	8.52	0.548				
No colon in title	566	8.740	69.80	8.355	0.3512	−2.389	201	−2.855	.004761
Colon in title	139	11.130	80.23	8.896	0.7597				

Note. The "count" is the number of papers in each subset, "mean" is the mean citation count for that subset, "mean diff" is the mean difference of the citation means between the two subsets, and *P*(*t*) is the student *t* test probability, where often small values (often <0.05) are used as an indicator of significance.

variance, though (e.g., papers with similar CR can have widely different citations). Therefore, it is best to think of this result in terms of the likelihood of higher citation increases with increasing CR.

3.2. Characteristics of Authors

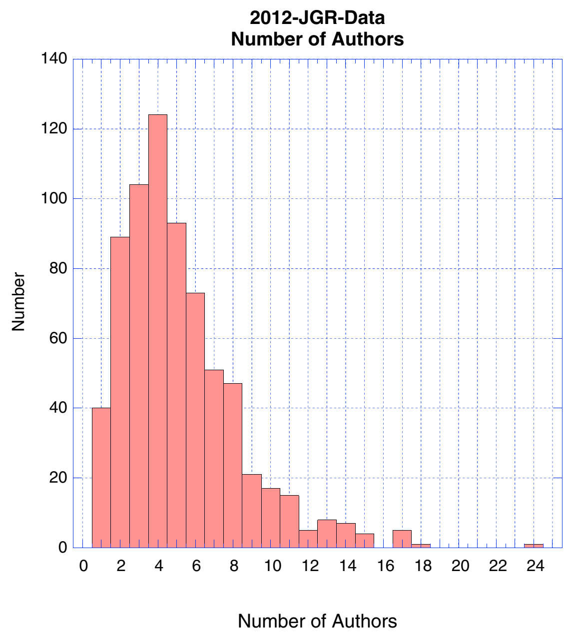
3.2.1. Number of Authors

There is a significant difference in the number of citations that a paper attracts and the number of authors. Figure 3 (top) shows the distribution of author number for the 705 papers, and Figure 3 (bottom) shows the correlation of author number and citation count. A Welch's Student *t* test shows that there is a significance difference between papers ($P(t) = 0.001$) that have above the median number of authors (five or more) compared to those with a smaller number of authors (Table 1). This small but statistically significant increase in citations with the number of authors is similar to the finding of Puuska et al. (2014) for the natural science category in their study. Previous studies have suggested two potential reasons for this difference—the first being that more authors allow more self-citations and a larger scientific network of colleagues, and the second potential reason is that there is an inherent "impact" difference between interdisciplinary team-based research and solo author or small group research (e.g., Fox et al., 2016). WoS includes a citation count that removes self-citations, but this only includes citations from papers by the same first author. That is, when calculating this metric, it does not take into account the self-citation influence of the coauthors. Therefore, it is difficult to determine the influence of this potential reason. The second hypothesis about why papers with more authors have more citations (that the additional authors make for a better paper) also cannot be readily tested with the available data. Perhaps the number of revision rounds could be compared, or the number of edits per manuscript page, but such numbers are often not tracked, so compiling these metrics would be extremely burdensome. It could be, and probably is, that both of these reasons contribute to the citation increase.

3.2.2. Role of Multi-Institution Teams and International Collaborations

We also examine if there is a citation advantage of papers with coauthors from a number of institutions. We further attempt to examine if there is a difference between international teams of authors compared to collaboration teams from a single country.

Similarly to what was found for total number of authors, there is an advantage in the mean number of citations for papers that have more than the median (four or more) institutions represented in the author list. The impact is smaller than for total authors, but the distributions of citations between those with 3 and fewer institutions represented in the author team compared to 4 or more institutions are statistically different.



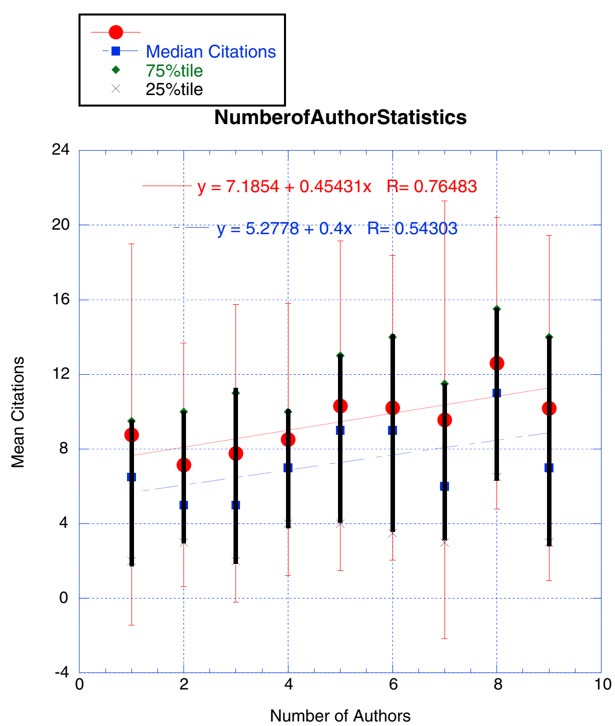


Figure 3. (a) The distribution of the number of authors per paper for our sample. (b) shows the distribution divided into deciles and the mean (with standard deviation) and median (with quartiles) of citations for each bin.

We also found that papers that had more countries represented in the author list follow the same trend as number of institutions. Just more than half of the papers had 2 or more countries represented in the author list (the median was 2), so the "above median" subset was for 3 or more countries in the affiliation list. We found that papers with more international coauthors garnered an average of 2.3 more citations than those with only one or two countries represented in the author list. Again, this is in line with the results from the natural sciences category of the Puuska et al. (2014) study, who found a large increase in the field normalized citation score for international teams versus domestic teams (note that their study considered Finnish-authored papers from 1990 to 2008).

3.2.3. Country of First Author

The citation statistics based on the country of the institution of the first author (not the nationality or country of origin of the first author) were examined, and the results for the top 11 countries in terms of total number of papers are presented in Table 2. Institutions in the USA are by far the largest contributor of papers to JGR-Space Physics in 2012, with China and Japan being the next largest producers of papers. The table shows the number of papers from each country that had eight or greater number of citations (i.e., just above the median, for a total of 332 or 47% of the total number of papers) to represent the top "half" and bottom "half" of the citation distribution. If the papers were randomly distributed, the percentage of papers from each country in the top "half" should be 47% (plus or minus square root of n [number of papers from country]). The last two columns examine the number and percentage of papers from each country that have 12 or more citations (roughly top 28% of papers) and those with 3 or less citations (bottom 28% of papers to examine if the distribution of papers in roughly the top and bottom quartiles). If the percentage is significantly above or below 28%, then the country is either over or under represented in the top or bottom quartile. Several countries' impact are below expectations (Japan, Canada, and India). However, comparing distributions of citation statistics for all papers published by each country, only Japan and India have statistically significant (>99% confidence interval) differences with papers published by U.S. institutions. Japan's mean difference with the USA is 3.019 citations, while India's is 5.239 citations. Canada has a mean citation difference with the United States of 2.99 citations, but the significance is weak (82% confidence interval or t probability of 0.18).

An examination of the citations distributions of papers from U.S. institutions compared to all papers from the rest of the world (Table 2) shows that there is a statistically significant (98.7% confidence level) but small (mean difference of the distributions of 1.6 citations) citation difference.

3.2.4. Gender of First Author

Of the 705 papers published in JGR-Space in 2012, we were able to identify the gender of 653 first authors. The 52 authors that were not gender identified (7.4% of the total) are excluded in the following analysis. Men constituted 81% (528) of the first authors, while women constituted 19% (124). This gender ratio is similar to the demographic survey results that found 83% of the Solar and Space Physics field were men and 17% were women (NRC, 2012). This indicates that unlike other disciplines, men and women publish first-author papers proportionally to their representation in the field (e.g., men do not publish more than women).

Table 2
Citation Differences among Papers from Different First-Author Country/Regions

Country/ Region	# of papers	# of papers with >8 citations	% of papers with >8 citations	Papers in bottom 28% tile	Papers in the top 28% tile
USA	321	172	53.6	76 (23.4%)	101 (31.5%)
PRC	64	34	53.1	11 (17%)	21 (33%)
Japan	61	20	32.8	26 (43%)	10 (16%)
UK	38	20	52.6	5 (13%)	9 (24%)
Canada	26	7	26.9	15 (58%)	5 (19%)
Germany	22	10	45.5	5 (23%)	7 (32%)
France	19	9	47.4	4 (21%)	6 (32%)
Russia	19	9	47.4	4 (21%)	6 (32%)
India	18	4	22.2	9 (50%)	3 (17%)
Taiwan	16	6	37.5	5 (31%)	3 (19%)
South Korea	16	6	37.5	5 (31%)	2 (12.5%)

The gender distribution of papers that appeared in the top quartile of citations (those with 12–69 citations) was 133 men, 26 women, and 8 unidentified. In this high-citation sample, the percentage of male first authors of those gender identified is 83.6%, while female first authors made up 16.4%. For the bottom quartile of papers (those with 0–3 citations), the distribution was 141 male, 37 female, and 16 unknown. This is 79.2% male and 21% female of the known gender authors. These numbers are consistent with a random distribution of papers.

The distributions of citations based on gender are statistically identical using a Welch's student t test with the mean citation rate for women 9.16 ± 8.73 with a median of 7. For men the mean is 9.46 ± 8.72 with a median of 7. Table 1 lists the statistical results of this comparison.

A recent study by Caplar et al. (2017) examined over 200,000 publications from 1950 to 2015 in astronomy and found a significant gender bias, with papers written by female astronomer first authors having about 10% less citations than male first-author papers. They attempted to separate other nongender determinants (e.g., seniority) and found that, though the gender difference has been decreasing over time, the bias still persists. Within the much more limited sample size of this study, the bias seen in astronomy is not seen in a statistically significant manner in space physics.

3.3. Word Analysis of Titles

Though titles are "less" important for discoverability in the age of electronic full-text search algorithms, they are still important for helping scientists identify useful papers when confronted with a large number of returned papers following a keyword search. For example, searching for the keywords "chorus" and "magnetosphere" for papers published in 2012 on WoS returns 56 papers (including 26 published in JGR, 10 in the AGU Monograph [Dynamics of the Earth's Radiation Belts and Inner Magnetosphere], 7 in GRL, and the remainder in 6 other journals). Is there a characteristic of the title that is correlated with its citation count? To address this question, we examined a number of characteristics of the title.

3.3.1. Number of Words in Title

We examined if the length of the title was predictive of the number of citations that a paper received. That is, is there a difference between short and long titles? There is not. The mean and median citation counts for the quartiles of the distribution had little variation and no statistically significant trends. We divided the papers into roughly quartiles based on the number of words in the title (titles with 3–10, 11–13, 14–16, and 17 to 30 words) and examined the mean and median of the citation count in each of the quartiles. The citation count distribution statistics (mean and median) were very similar between the bins with no consistent trend (the top and bottom quartiles had statistically identical distributions according to a Welch's student t test).

Other studies divided papers into long and short titles by using the median of the number of words to divide the distribution (Paiva et al., 2012). Looking at papers with long titles (those with the number of words exceeding the median word count of 13 words) compared with short titles (those with the number of words less than or equal to the median), we found a small and statistically insignificant difference, with shorter titled papers having a mean citation count of 9.46 ± 9.12, while long titled papers have a mean citation count of 8.93 ± 7.82. A Welch's student t test found that these distributions are statistically different at only a 40% confidence level. This result is marginally consistent with some studies that found a correlation between short titles and higher citations counts (Letchford et al., 2015; Paiva et al., 2012), though in our study, the trend is not significant.

3.3.2. Acronyms in Title

Discouraged by AGU's Author Guide (http://publications.agu.org/author-resource-center/text-requirements/) and one of the authors of this study (https://liemohnjgrspace.wordpress.com/2014/05/08/acronyms-in-paper-titles/) is the inclusion of acronyms and abbreviations in the title. However, some studies have found that the use of acronyms in the title increases the citation count (e.g., Jacques & Sebire, 2010). An examination of the papers published in JGR-Space Physics in 2012 found that most (463 or 65.7%) papers did not have an acronym in the title. There were 163 (23.1%) with 1 acronym and 79 (11.2%) with 2 or more acronyms (with

Table 3
Number of Papers with a Colon in Title in each Citation Count Quartile

	Number of papers with a colon in the title	% of papers in quartile bin with a colon in the title
0–175 (top quartile)	48	27.3% (+/−) 2.3%
176–352	40	22.7% (+/−) 3.6%
353–528	29	16.6% (+/−) 3.1%
529–705 (bottom quartile)	22	12.4% (+/−) 2.5%

Note. The error given is proportional to the square root of *N* (number of papers in bin).

one paper having five acronyms or abbreviations in the title "Global S4 index variations observed using FORMOSAT-3/COSMIC GPS RO technique during a solar minimum year"). It should be noted that many acronyms included in titles were names of satellites. We found there was no statistical difference of the citation count between the papers with no, one, or two or more acronyms with the distributions of citation counts for papers with acronyms and those without essentially identical according to a Welch's student *t* test of significance (see Table 1).

3.3.3. Colon in Title

The use of a colon to make a compound title is relatively common, and some studies suggest that it is correlated with higher citation rates (e.g., van Wesel et al., 2014), while a study of six PLoS journals found an anticorrelation (Jamali & Nikzad, 2011).

For our study, we found that 139 (19.7%) of the papers contained a colon. We found that the citation mean and median count was higher for papers with a colon in the title. Papers with a colon had a mean citation count of 11.13 ± 8.96 and a median of 9 citations, while papers without a colon in the title had a mean citation count of 8.74 ± 8.36 and a median citation count of 6.5. A Welch's Student *t* test shows that the two distributions are statistically different above the 99% confidence level. Another way to look at the difference is that papers with titles that contained a colon were slightly overrepresented in the top half of the sample (they made up 24.9% of all papers that had citation counts above the median). Papers with colons were significantly underrepresented in the bottom half of the citation distribution (they made up 14.5% of the low-impact half of the papers). Table 3 shows the number of papers that contain a colon in the title in each of the four quartile bins of citation count. If the papers were randomly distributed, the percentage in the third column should be the same as the total fraction of papers with a colon compared to the entire population (19.7%). Papers with a colon are overrepresented in the upper quartile and underrepresented in the bottom quartile, suggesting that papers with a colon have a higher probability of being highly cited than those without colons.

3.3.4. Geophysical Region Names Used as Adjective or Noun in Title

An analysis of the words used in titles was conducted examining the most common words used in high-impact papers; if there are differences in citation counts if regions of the space environment are used as nouns (ionosphere, magnetosphere, and thermosphere) or as adjectives (ionospheric, magnetospheric, or thermospheric). The most common space environment region used in titles was the ionosphere, with 72 papers having "ionosphere" in the title and 76 papers with "ionospheric." Papers using "ionosphere" in the title had a mean citation count of 9.56 ± 8.378 (median of 7.5), while those with "ionospheric" had a mean citation count of 7.567 ± 6.14 (median of 6). The difference is significant at the 10% level (Welch's *t* test probability of 0.103). A search of WoS using the keyword "ionosphere" as the "Topic" found 270 papers within JGR in 2012. Using the keyword "ionospheric" found 216 papers. A similar citation difference trend was found for "magnetosphere" and "magnetospheric" (49 papers had "magnetosphere" in title with a mean citation count of 12.41 ± 10.60 and 22 papers with "magnetospheric" having a mean citation count of 11.05 ± 12.01), though the distributions are not significantly different (Welch's student *t* test probability of 0.65). Searching WoS using the keyword "magnetosphere" as the "Topic" search found 260 papers, while using "magnetospheric" found 113 papers. So if authors use a region as an adjective, the noun counterpart should be used in the keywords or abstract to enable greater discoverability, though the impact on citations is weak, if at all.

4. Conclusions, Caveats, and Discussion

There are several conclusions to be summarized from this study. For the JGR-Space Physics, the citation distribution follows a lognormal distribution (Figure 1) similar to what is found for other journals and subject areas. By analyzing characteristics of papers by median, quartile, and decile, we found there are significant differences between highly cited papers and papers with few citations—especially with regards to the number of CRs contained in the paper and the number of authors, institutions, and countries represented on the manuscript. We also found statistically significant differences for only one title characteristic—the inclusion of a colon. Compound titles that use a colon receive higher citations. We did not find any statistically significant difference for papers with long or short titles or those with acronyms. We did not find a significant difference in citations for papers with male or female first authors, in contrast to what has been found in other disciplines including astronomy. We found that most countries had identical citation statistics compared to papers with U.S. first-author institutions but that there was a statistically significant, but small citation advantage for U.S. institution papers compared to the rest of the world.

One issue with comparing subsets of papers with respect to citations is that they are drawn from a highly skewed nonnormal distribution (Figure 1). This makes statistics based on the distribution highly dependent on outliers and therefore can impact the Journal Impact Factor (e.g., Mutz & Daniel, 2012; Royle, 2015) and statistical studies of subsets of the data. Of particular concern is the interpretation of standard deviation; it cannot be assumed that 50% of values are below the mean and that 68% of the values fall within $\pm 1\sigma$ of the mean. For the citations presented in Figure 1, 62% of values fall below the mean. In addition, it indicates that correlation coefficients could be overly influenced by a few high-citation outliers.

Another issue of using citations as a measure of the "impact" is the complex reasons that papers are cited (e.g., Adler et al., 2008). Cozzens (1989) proposed that there are two main motivations to cite a paper—one is an acknowledgment of the intellectual contribution to the work and the other is a "rhetorical" reference describing previous work (and in the case of citing a review paper, the citation does not necessarily acknowledge the persons that actually did the underlying work). This type of citation is made to carry on the scientific conversation and not necessarily to acknowledge the intellectual contribution. Hence, caution is needed when interpreting citation statistics—especially when using them for assessment of the quality of the work.

However, we suggest that this study (albeit a study of only 1 year from one journal) provides several key insights on characteristics of high-citation papers and our community. The analysis suggests two steps that can be taken in the research phase and manuscript preparation stage to maximize the potential for more citations to the eventual publication. The first recommendation to the community is to cite the relevant literature. For papers with at or above the median CRs (39), there was an increase of 3.5 citations, on average, over papers with below the median number of CRs. The second piece of advice is that you should collaborate, especially internationally, on the research effort. At the granulation level considered in this study (5 or more compared to 4 or fewer), more authors lead to more citations, and the inclusion of authors from other countries helps even more. Remember that the variance in the citations for any particular metric is large. A paper that does not fit these recommendations could still be highly cited. Nevertheless, for each of these key factors, the increase in average citations is on the order of an extra citation per year over the interval considered by this study, the first ~4 years after publication. The study also found that there is no statistical difference in citation rates for most countries that publish meaningful number of papers, but there is a slight bias toward U.S. institution first-author papers compared to the rest of the world. We found that women publish at the same level as their representation in the field and there is no statistically significant difference between the citation distributions of first author female and male papers.

Acknowledgments
M.B.M. is a past editor-in-chief of AGU's Reviews of Geophysics (2010–2017) and editor of Geophysical Research Letters (2004–2006), and M.W.L. is currently the editor-in-chief of AGU's JGR-Space Physics. This study grew out of work done for the Solar and Space Physics Decadal Survey, which was orted by an NSF Rapid grant (NSF 1137659). The authors thank Jory Lerback (formally of AGU) for help with gathering Wiley usage data. The database of bibliometric records used in this study is available at the UM Deep Blue data archive (doi: 10.7302/Z21C1V2H). Yes, the colon in the title was explicitly included to conform to "best practice" for higher citations.

References

Acuna, D. E., Allesina, S., & Kording, K. P. (2012). Future impact: Predicting scientific success. *Nature*, *489*(7415), 201–202. https://doi.org/10.1038/489201a

Adler, R., Ewing, J., & Taylor, P. (2008). IMU report, citation statistics. Retrieved from http://www.mathunion.org/Publications/Report/CitationStatistics, Accessed June 1, 2017.

Bornmann, L., & Daniel, H.-D. (2007). What do we know about the h index? *Journal of the American Society for Information Science*, *58*(9), 1381–1385. https://doi.org/10.1002/asi.20609

Bornmann, L., & Daniel, H.-D. (2008). What do citation counts measure? A review of studies on citing behavior. *Journal of Documentation*, *64*(1), 45–80. https://doi.org/10.1108/00220410810844150

Caplar, N., Tacchella, S., & Birrer, S. (2017). Quantitative evaluation of gender bias in astronomical publications from citation counts. *Nature*, *1*, 0141. https://doi.org/10.1038/s41550-017-0141

Cozzens, S. E. (1989). What do citations count? The rhetoric—First model. *Scientometrics*, *15*(5-6), 437–447. https://doi.org/10.1007/BF02017064

Fox, C. W., Paine, C. E. T., & Sauterey, B. (2016). Citations increase with manuscript length, author number, and references cited in ecology journals. *Ecology and Evolution*, *6*(21), 7717–7726. https://doi.org/10.1002/ece3.2505

Garfield, E. (2006). The history and meaning of the journal impact factor. *Journal of the American Medical Association*, *295*(1), 90–93. https://doi.org/10.1001/jama.295.1.90

Haslam, N., Ban, L., Kaufmann, L., Loughnan, S., Peters, K., Whelan, J., & Wilson, S. (2008). What makes an article influential? Predicting impact in social and personality psychology. *Scientometrics*, *76*(1), 169–185. https://doi.org/10.1007/s11192-007-1892-8

Haustein, S., Costas, R., & Larivière, V. (2015). Characterizing social media metrics of scholarly papers: The effect of document properties and collaboration patterns. *PLoS One 10*(3), e0120495. https://doi.org/10.1371/journal.pone.0120495. pmid: 25780916

Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(46), 16,569–16,572.
Jacques, T. S., & Sebire, N. J. (2010). The impact of article titles on citation hits: An analysis of general and specialist medical journals. *JRSM Short Report*, *1*(1), 2–5. https://doi.org/10.1258/shorts.2009.100020
Jamali, H., & Nikzad, M. (2011). Article title type and its relation with the number of downloads and citations. *Scientometrics*, *88*(2), 653–661. https://doi.org/10.1007/s11192-011-0412-z
Letchford, A., Moat, H. S., & Preis, T. (2015). The advantage of short paper titles. *Royal Society Open Science*, *2*(8), 150266. https://doi.org/10.1098/rsos.150266
Mukherjee, S., & Romero, D. M. (2017). Ben Jones and Brian Uzzi. The nearly universal link between the age of past knowledge and tomorrow's breakthroughs in science and technology: The hotspot. *Science Advances*, *3*(4), e1601315. https://doi.org/10.1126/sciadv.1601315
Mutz, R., & Daniel, H. D. (2012). Skewed citation distributions and bias factors: Solutions to two core problems with the journal impact factor. *Journal of Informetrics*, *6*(2), 169–176. https://doi.org/10.1016/j.joi.2011.12.006
National Research Council (2012). *Solar and Space Physics: A Science for a Technological Society*. Washington, DC: The National Academies Press.
Paiva, C. E., Lima, J. P. d. S. N., & Paiva, B. S. R. (2012). Articles with short titles describing the results are cited more often. *Clinics*, *67*(5), 509–513. https://doi.org/10.6061/clinics/2012(05)17
Penas, C. S., & Willett, P. (2006). Brief communication: Gender differences in publication and citation counts in librarianship and information science research. *Journal of Information Science*, *32*(5), 480–485. https://doi.org/10.1177/0165551506066058
Puuska, H. M., Muhonen, R., & Leino, Y. (2014). International and domestic co-publishing and their citation impact in different disciplines. *Scientometrics*, *98*(2), 823–839. https://doi.org/10.1007/s11192-013-1181-7
Royle, S. (2015). Wrong number: A closer look at impact factors. Quantixed. Retrieved from https://quantixed.wordpress.com/2015/05/05/wrong-number-a-closer-look-at-impact-factors/, Accessed 18 May 2017.
San Francisco Declaration on Research Assessment (2012). Retrieved from http://www.ascb.org/dora/
Sin, S.-C. J. (2011). International coauthorship and citation impact: A bibliometric study of six LIS journals, 1980–2008. *Journal of the American Society for Information Science and Technology*, *62*(9), 1770–1783. https://doi.org/10.1002/asi.21572
Tang, X., Gong, W., Yuan, F., Li, R., Han, X., Huang, S., et al. (2016). Top-cited articles in digestive system disease from 1950 to 2013. *Journal of Gastroenterology and Hepatology*, *31*(1), 107–111. https://doi.org/10.1111/jgh.13032
Van den Besselaar, P., & Sandström, U. (2016). Gender differences in research performance and its impact on careers: A longitudinal case study. *Scientometrics*, *106*(1), 143–162. https://doi.org/10.1007/s11192-015-1775-3
van Wesel, M., Wyatt, S., & ten Haaf, J. (2014). What a difference a colon makes: How superficial factors influence subsequent citation. *Journal of Scientometrics*, *98*(3), 1601–1615. https://doi.org/10.1007/s11192-013-1154-x
Waltman, L., & van Eck, N. J. (2013). Source normalized indicators of citation impact: An overview of different approaches and an empirical comparison. *Scientometrics*, *96*(3), 699–716. https://doi.org/10.1007/s11192-012-0913-4

Erratum

The heading of column 1 in Table 2 has been updated to better describe the contents of column 1, and the present version may be considered the authoritative version of record.