Check for
updates

Citations and certainty: a new interpretation of citation counts

Henry Small[1] · Kevin W. Boyack[2] · Richard Klavans[3]

Abstract

We report that the rate of hedging in citing sentences for biomedical papers is inversely related to the citations received by the papers as measured by the number of citances in citing papers. Hedging is often regarded as an expression of uncertainty in rhetorical studies of scientific text. Citing sentences, or citances, are retrieved from the PubMed Central database for papers having 10 or more citances, and the percentage of citances containing hedging words is plotted against the number of citances for the papers, which is closely related to the citation count. Hedging rates are computed separately for method and non-method papers, the latter being more frequently hedged. Rates of hedging are found to be higher for papers with fewer citances, suggesting that the certainty of scientific results is directly related to citation frequency. Similarly, early citations made soon after publication are more hedged than later citations. The implications of this finding for the interpretation of citation counts are discussed, and the directions for future research.

Keywords Citations · Citances · Hedging · Uncertainty · Biomedicine

Introduction

The interpretation of citation counts has been a perennial problem in the field of scientometrics and informetrics from its earliest years. High citation counts have been variously interpreted as indicators of peer recognition, importance, quality, impact, and influence. Such interpretations have been used to justify the use of citation counts in the evaluation of scholarly output, and numerous formulas have been deployed to compensate for the

✉ Henry Small
  hsmall@mapofscience.com

  Kevin W. Boyack
  kboyack@mapofscience.com

  Richard Klavans
  rklavans@mapofscience.com

[1] SciTech Strategies, Inc., Bala Cynwyd, PA 19004, USA

[2] SciTech Strategies, Inc., Albuquerque, NM 87122, USA

[3] SciTech Strategies, Inc., Wayne, PA 19087, USA

myriad confounding factors such as field of science, language, nationality, age and type of paper, and journal of publication, etc.

In their pioneering book Jonathan and Stephen Cole argued that citations can be interpreted as the "quality" of scientific work (Cole and Cole 1973). This was based mainly on the correlation of citation counts with expert opinions or honorific awards such as Nobel Prizes. Other studies pointed to the correlation between citation counts for scientists at a university with survey-based university rankings (Narin 1976). Yet another approach used expert opinions on the "importance" of individual papers compared to their citations. Virgo asked experts to select which paper of a topically matched pair was the more important (Virgo 1977), and Small (1977) asked experts to name important developments in a specialty and the papers responsible for them. While these paper-based studies are more direct than the studies based on correlations, the problem is that the terms "quality" and "importance" remain undefined. We do not know what factors motivated the experts in making these judgments or the motivations of the authors citing these papers. Also unexplored is whether other descriptions such as "useful" or "valid" would have yielded similar results. In constructivist theories, citations are not motivated by any intrinsic quality of the cited work, but rather the ability of the reference to bolster the validity of the authors' knowledge claim (Latour and Woolgar 1979) and persuade the reader (Gilbert 1977).

A comprehensive review of the literature on "what citation counts measure" was undertaken by Bornmann and Daniel (2008). This review deals primarily with citer motivation studies based on the notion that if we understand why people cite we will understand what citation counts mean. The authors conclude that the traditional normative view of citations as a measure of communal recognition is justified on a macro- but not necessarily a micro-level, where constructivist forces prevail. However, none of the motivational factors examined seem capable of yielding a consistent account of why some papers are cited more than others. How can we accept aggregate citation counts as valid indicators of peer recognition if they are the result of multiple idiosyncratic decisions at the micro-level? Are we to conclude that citation differences are only the result of random micro-effects that confer advantage or disadvantage? Are the causes of citation differences too complex for us to untangle? A clue is provided in the last sentence of the review that, following Cole (1992), perhaps citations indicate what "knowledge is accepted as given" (Bornmann and Daniel 2008).

Another clue is provided by lists of most cited papers which are dominated by method papers (Garfield 1990; Van Noorden et al. 2014). Methods appear to be mundane and boring, but often provide the factual basis for scientific knowledge. Some have argued that the solution to this puzzle may lie in the rhetorical analysis of citation contexts. Indeed, Latour argued early on that the "game" of science was to push statements towards the status of "facts" by rhetorical maneuvers in scientific texts (Latour and Woolgar 1979). In rhetorical studies of scientific text, hedging is interpreted as an expression of uncertainty (Chen and Song 2018), the inverse of certainty. Hyland has pioneered in the study of hedging which he defines as ". . . devices which withhold complete commitment to a proposition . . . ", and that are based on " . . . plausible reasoning rather than certain knowledge . . ." (Hyland 2009, 75) More recently, Chen and Song have used deep learning methods to identify a wide range of hedging words and phrases in scientific texts with the objective of finding gaps in biomedical knowledge (Chen and Song 2018). The sentences containing references in scientific texts were found to contain a higher concentration of hedging terms than other parts of the text (DiMarco et al. 2006). Recently Small (2018) observed that the citation contexts for lower cited papers contained more occurrences of the hedging word "may"

than more highly cited papers, controlling for method papers which generally were associated with less hedging.

The purpose of the present report is to extend this analysis to lower citation levels, expand the set of hedging words, and show how hedging relates to the age of the reference. We will discuss "certainty" as a new way to interpret citation counts, how earlier interpretations might be affected, and how this finding raises new questions for citation analysis and evaluation.

Methods and data

To study the relationship between citation frequency and hedging terms in citation contexts, it is necessary to use a large corpus of scientific papers in full text from which citing passages can be extracted along with their associated cited references. The number of times specific cited references are mentioned also needs to be determined. The data source used in this study was the open access subset from PubMed Central (PMC) updated to October 1, 2017. The procedure for converting this full text XML data set to a relational data base was described previously (Small 2018). Although papers in PMC cite previous work from many different disciplines, our results are limited to cited papers for which a PubMed ID (PMID) appears in the XML reference lists. As explained in that paper, each citing sentence, called a "citance" (Nakov et al. 2004), was extracted and linked to the specific reference cited. Through use of a MySQL relational data base, it was possible to identify all the citances associated with a given cited paper and whether those citances contained specific hedging words.

In this paper we use the number of citances, or citing sentences, for a given cited paper instead of the traditional number of citations which only counts each citing paper once. The number of citances (or "mentions" as it is also called) weights each traditional citation by the number of "mentions" or repeats of each reference within the citing texts. The citance count is higher than the traditional citation count by about 57% (Boyack et al. 2018). For example, if paper X is mentioned two times in paper A and three times in paper B, then X has a citance count of five, while its traditional citation count is two. Since we are concerned with hedging words that occur within the citing sentences, which can vary within a given paper, it seems appropriate to base the analysis on the number of citances rather than the number of traditional citations, although the two types of counts are highly correlated with a coefficient of 0.97 across 1 million papers with 10 or more citances from PubMed Central. Thus, in general, a paper with a high citance count will also have a high traditional citation count, and vice versa.

To study how hedging is related to citance frequency, a citance count was determined for each cited paper having a PubMed identifier in the PMC data base. Then all cited papers were taken across the various ranges of citance counts from the highest count of 13,936 for one paper down to a count of ten citances per cited paper. Below the level of 10 citances per paper, samples were taken of papers with one and five citances.

A very limited set of three hedging terms was used: "may", "could", and "might". The rationale for using such a limited set will be discussed below. Citances were searched for the presence of hedging word strings in MySQL that included spaces before and after the words, as in '% could %', which retrieves instances in which the hedging word was embedded in the sentence. To avoid selecting proper nouns such as "May", a binary regular

expression was used, as in REGEXP BINARY 'may'. No part-of-speech tagging was performed due to the unambiguous nature of these words.

Two approaches can be taken to determine the rate of hedging within a sample of citances: computing the percentage of citances containing hedging terms, what could be called the "per citance rate", and the percentage of hedging words of total words in the citances, which could be called the "per word rate". The primary method used in this study is the "per citance rate", but the "per word rate" is also used to check the former since the "per word" approach can take into account differences in citance lengths. The statistical test employed for comparisons of "per word rates" is log likelihood as implemented in the Wordsmith Tools software (Scott 2004).

To test the statistical significance of differences in "per citance rates", the difference in average hedging rates for cited papers can be used, or alternatively, the difference of aggregate hedging rates computed by combining the counts for all cited papers. Because the distribution of hedging is not normal, non-parametric tests were used such as Chi square for aggregate hedging rates or the Wilcoxon–Mann–Whitney test for cited paper averages.

The plan of the paper is as follows. We first describe the distinctive hedging behavior shown by method papers, and how we can distinguish them from other papers. We then show results for the most common hedging word "may" and plot the hedging rate against a rank ordering by number of citances. An approach to expanding the set of hedging words is described, giving examples of citances containing those terms. The effect of the expanded hedging word set on the plot of hedging rate versus rank ordering by number of citances is discussed, followed by the transformation obtained by logging the number of citances rather than rank ordering them. Statistical tests are applied based on specified intervals of citance frequency. The relationship of hedging and citance age is discussed, as well as a corroboration of the relationship between hedging and citance counts using a corpus linguistics method at the word level.

Results

Hedging of methods versus non-method papers

In a prior study of the 1000 most cited papers from PubMed Central (Small 2018), it was found that citations to method papers were not hedged with the word "may" as often as non-methods, suggesting that method papers as a class should be examined separately from other papers, which we call non-method papers. Furthermore, it was found that lesser cited papers were hedged more often than more highly cited papers regardless of whether a paper was a method or not. The category of non-method papers includes items such as discovery papers, reviews, and regular research articles. Across only the top 1000 papers, however, the trend toward higher hedging for lower cited papers was not very marked: from the most cited to least cited paper, the increase in hedging rate was only about 1% point, 0.5% to 1.5% for methods, and 3.5% to 4.5% for non-methods. The hedging rate was measured by counting the percentage of citances for each cited paper that contained the word "may", and then averaging those percentages for sets of 100 papers. Alternatively, we summed the hedged citances across a set of 100 papers and divided by the total citances to get a rate for the group.

Based on these initial findings, it was decided to look beyond the top 1000, down to papers having 10 citances, comprising over 1 million papers, to see if the hedging rate

would continue to increase for lesser cited papers. To control for the lower hedging rate of methods, and the possibility that they might mask the effect of the more strongly hedged non-methods, a logistic model was developed to differentiate methods from non-methods at these lower levels. This was based on the finding that methods could be identified by looking for "utility" words ("using" or "used") in their citances, and by computing the percentage of citances that appear in citing papers' "method" sections (Small 2018). A combination of these criteria was used to classify the lesser cited papers down to the ten citances per paper level. A paper was classified as a method if either the logistic regression equation for "utility" words or the equation for the "method" section variable gave a probability greater than one-half, using method or non-method as the dependent variable.[1] The accuracy of this assignment was 91.2% based on a manual classification of 1000 papers.

The logistic model predicted that 18% of papers in the set of 1 million papers with 10 or more citances were methods, although this percentage varied over the citance frequency range from a high of 50% methods for the top 1000. At lower citance frequencies, the rate was around 18% with a slight increase in rate at the lowest citance frequencies from its minimum cumulative rate of 17.2% (see Fig. 1).

As noted above, the previous study was undertaken with just the single hedging word "may". When this approach is extended beyond the top 1000 papers by number of citances, the result is that method and non-method papers show a gradual increase in the percentage of hedging. Figure 2 shows this trend for the 1,025,496 papers with 10 or more citances. Across this range, non-methods are about 2.3% points higher in hedging rate than methods, and the increase in hedging across the first 10,000 sample is 2.8% points. The largest gains in hedging occur at the left of the curve (see Fig. 2) for roughly the most cited 50,000 papers. After that point, moving to the right toward lower citance counts, the rate of hedging gradually increases by about 0.12% per 100,000 papers. At the same time the citance count deceases by about four orders of magnitude, thus clearly showing the increase in hedging with decreasing citance count. One objective of this study is to investigate the relationship between the citance count and hedging rate.

Expansion of the hedging word set

Restricting the analysis to the single hedging word "may", even though it was the most commonly used hedging term, seemed an unnecessary limitation given the wide range of hedging words or phrases that have been identified (Chen and Song 2018). The approach taken to expanding the hedging word set was to use a lexicon of hedging terms from a previous study, compiled from various sources (Small and Klavans 2011), and then determine which of these words was likely to have the greatest impact on the analysis. The list of 52 previously compiled hedging terms was matched against two samples of citances. One sample focused on highly cited method papers having greater than 1600 citances and were expected to contain few hedging terms, while the other sample focused on low cited, non-method papers having 12 citances which were expected to contain more hedging terms. The samples consisted of roughly the same number of citances.

---

[1] The logistic regression equations used were $p1 = 1/(1 + \mathrm{Exp}(3.424 - 0.216 * [\% \text{ utility}]))$ where [% utility] is the percentage of citances containing utility words, and $p2 = 1/(1 + \mathrm{Exp}(2.72 - 0.11 * [\% \text{ method section}]))$ where [% method section] is the percentage of citances appearing in method sections. When either probability p1 or p2 is > 0.5, the paper was designated a method. All other papers were designated non-methods (Small 2018).
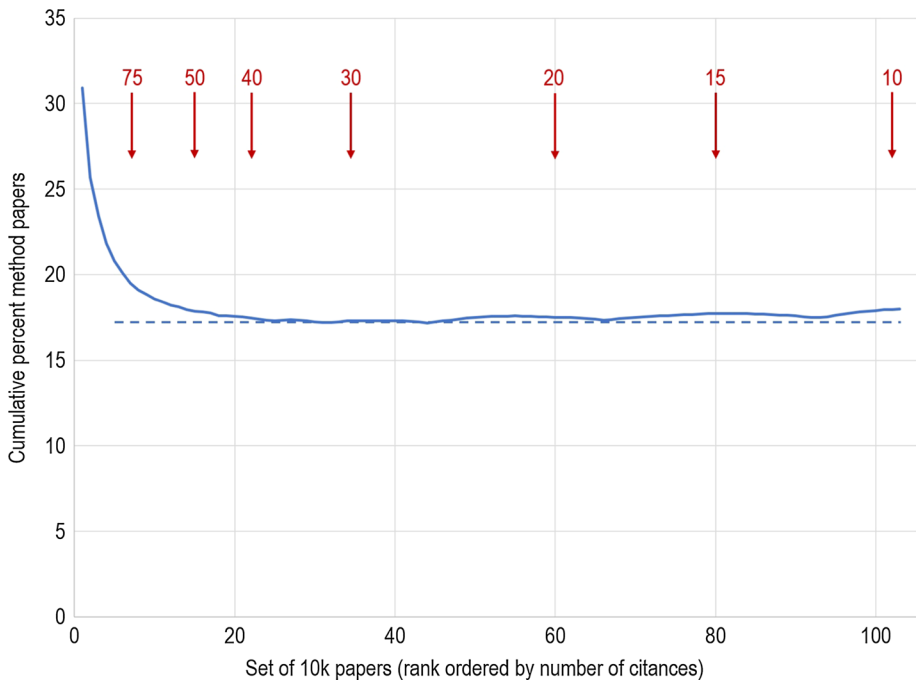
Fig. 1 Cumulative percent of method papers among papers with at least 10 citances grouped in sets of 10,000. Arrows show corresponding number of citances. The dotted line indicates the minimum cumulative rate of 17.2%. Numbers of methods are cumulated from highest to lowest number of citances

Table 1 presents the 10 hedging terms from the lexicon that occurred most frequently in the more highly hedged sample. The table gives the number of citances containing the specific term in the highly hedged sample, the number of citances in the low hedged sample, the percentage of total citances for the cited papers for each sample, and the ratio of low to high certainty percentages. For example, "may" has the highest occurrence in the low certainty sample and the highest percentage of total citances, and a much lower frequency and percentage in the high certainty sample. The second ranking word was "could", which is followed by "potential" and "might". The word "potential", in addition to denoting "possibility", is also used to describe electrical phenomena, and, to avoid introducing noise, it was excluded from the expanded hedging set. The expanded set thus consists of the three words "may", "could" and "might", three of the so-called modal auxiliaries.

Examples of hedging citances

Our underlying assumption is that the appearance of a hedging term in a sentence containing a reference confers uncertainty on the specific reference cited, or more precisely on ideas associated with the reference. Furthermore, the higher the fraction of citances containing hedging terms, the higher the degree of uncertainty of the idea or paper. Requiring the hedging term to occur in the same sentence as the reference increases the likelihood that the associated content is regarded as uncertain by the citing author.
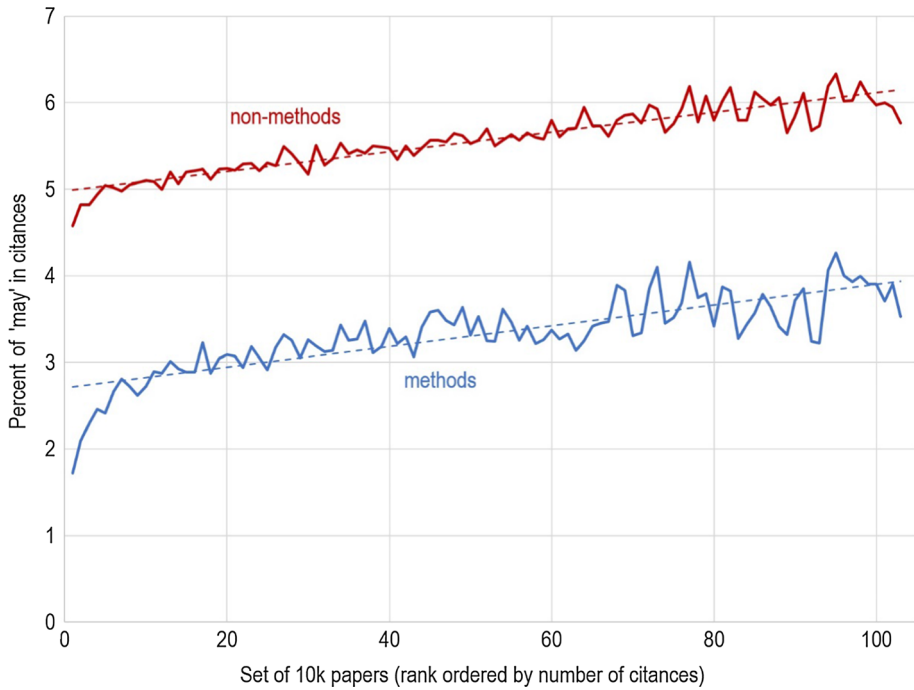
Fig. 2 Percentage of "may" citances for method and non-method papers by set of 10,000 papers ordered by total citances. Percentages of "may" citances are averaged for each set. Dotted lines show linear least-square fits for each curve

Table 1 Comparison of high and low certainty citance samples. For each of the ten most frequently occurring hedging terms in the low certainty sample, the table gives the number and percentage of citances in the low and high certainty samples, and the ratio of the low to high percentages

Word	Low certainty citances	High certainty citances	Percent low certainty	Percent high certainty	Ratio low to high
May	24,121	2708	12.72	0.70	18.14
Could	9005	1833	10.03	0.47	21.38
Potential	7042	3458	10.36	0.89	11.71
Might	4852	622	9.48	0.17	55.14
Likely	4449	1549	9.67	0.43	22.56
Possible	3530	2571	9.24	0.66	13.98
Whether	2576	1772	9.03	0.48	18.89
Appears	1885	118	8.86	0.07	131.52
Potentially	1689	502	9.04	0.15	58.79
Seems	1475	88	8.86	0.06	141.03

This effect can be illustrated by examination of examples of hedging citances in Table 2 where three citances have been selected for each of three papers on different topics. For example, in the case of the first paper on so-called phantom pain, the explanation given in

Table 2 Examples of hedged citances for three papers. For three cited papers (numbered 1 through 3 in the table), three citing sentences or citances are provided, each containing one of the three hedging words "may", "could" and "might"

---

1. Phantom pain: Harris, J. 1999. Cortical origin of pathological pain. The Lancet, 354(9188), 1464–1466

It has been hypothesized that it **may** be due to efferent motor cortical activity continuing without afferent sensory feedback to dampen motor commands (Harris, REF)

Some authors have reported that neuropathic pain after amputation **could** be due to the mismatching between motor output and sensory feedback (Mayer et al., REF), in agreement with the model of pathological pain by Harris (REF)

A mismatch between the brain's motor control and sensory systems has been suggested as one mechanism whereby maladaptive neuroplastic changes **might** contribute to the experience of chronic pain (Harris, REF; McCabe, REF)

2. Blood clotting: Andrews, D.A. & Low, P.S. 1999. Role of red blood cells in thrombosis. Current Opinion in Hematology, 6(2), 76–82

Moreover, eryptosis **may** affect the microcirculation, as PS-exposing erythrocytes may bind to the vascular wall and participate in blood clotting (REF, REF)

Furthermore, there is emerging evidence that RBCs **might** even play an active role in the process of thrombus formation, and Ca^{2+} appears to be central for this process as well REF, REF, REF)

At least in theory, erythrocyte adhesion to the vascular wall **could** compromise microcirculation and thus interfere with blood flow (REF, REF, REF, REF, REF, REF)

3. Cancer stem cells: Lu, J., et al. 2005. MicroRNA expression profiles classify human cancers. Nature, 435 (9), 834–838

In recent years, considerable interest has developed in the possibility that cancer stem-like cells (CSCs) in human tumors **could** be major contributors to resistance of tumors to conventional photon radiotherapy (RT) (REF, REF)

Similar mechanisms **might** explain therapeutic resistance of tumor stem-like cells towards both chemotherapy (REF) and radiation (REF, REF)

The resistance of CSCs to conventional chemo-radiotherapy **may** involve enhanced DNA damage repair pathways and alterations in cell cycle kinetics (REF)

---

the original paper cited is regarded as unsettled. In the second paper on blood clotting, red blood cell adherence to the vascular wall is seen as a possible but not certain mechanism. From these examples, uncertainty is being expressed regarding specific concepts associated with one or more of the references cited. It is important to note, however, that hedging does not assert that the paper is wrong, but only suggests that uncertainty surrounds some aspect of the ideas put forward. Papers can increase or decrease in their hedging rate over time as their knowledge claims are evaluated. An increase can signify that a paper initially seen as unproblematic, is now regarded as raising new questions. A good example of this is the third paper in Table 2 on cancer stem cells which went from a hedging rate of 5% 4 years after publication to 15% after 10 years, reflecting new questions regarding the clinical utility of miRNA expression monitoring in cancer treatment. However, most papers, as we will see, have decreasing hedging rates over time. Absence of hedging, of course, does not mean that a paper will avoid becoming obsolete, obliterated, and eventually uncited.

Hedging at lower citation levels

Obviously searching citances with an expanded hedging word set will increase the number and percentage of citances containing hedging words. This is clear from Fig. 3 which plots the three-word hedging rates of method and non-method papers for the 1,025,496
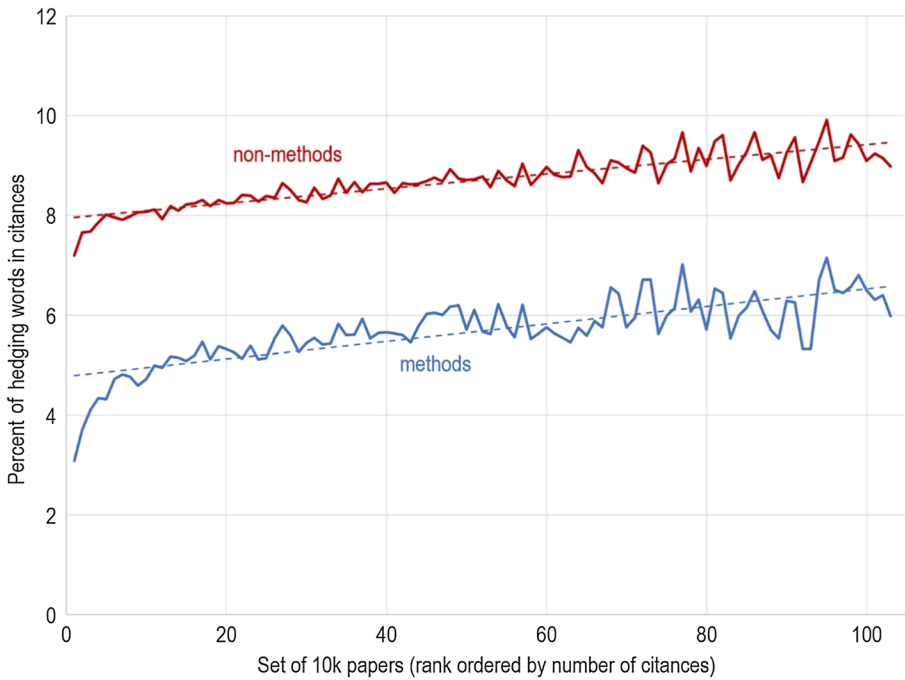
Fig. 3 Percentage of three-word hedging citances for method and non-method papers by set of 10,000 papers ordered by total citances. Percentages of three-word hedging citances are averaged for each set. Dotted lines show linear least-squares fits for each curve

papers with 10 or more citances. The curve is similar in shape to Fig. 2 for the single hedging word "may". In Fig. 3, however, the curves for the expanded hedging set are displaced upwards by about 3.2% points for non-methods and 2.4 points for methods relative to Fig. 2, while the range from most to least cited in Fig. 3 is about 2.5% points for non-methods and methods, slightly larger than was seen for "may" alone in Fig. 2.

The distribution of three-word hedging rates over the full set of over one million papers having 10 or more citances is non-normal and shifted to the low end. Two-thirds of the papers have a hedging rate < 10% and a third of papers are at or above 10%. One-fifth of papers have 0 hedging. The distribution, which is not shown, has a mean of 8.2%, with 42% of hedging rates above the mean and 58% below. A long tail extends to the right, and 1.4% of papers have a hedging rate of 30% or more. The highest observed rate is 80% for two non-method papers with 10 citances each.

The papers with < 10% hedging have a mean number of citances of 37.7 while papers with 70% or greater hedging have a mean of 11.1. Thus, there is clearly a trend towards increased hedging for papers with lower numbers of citances.

To plot the trend down to papers with 10 citances, we adopt a base ten logarithmic scale doubling the number of intervals to show more detail (Fig. 4). Each interval corresponds to a range of citance frequencies given on the x-axis. Logging the citance counts effectively linearizes the plots of hedging rates, and each order of magnitude increase in citances results in roughly a 2% decrease in hedging rate, with a slight falling off for non-methods at low citances. This means that the number of citances has an approximate inverse
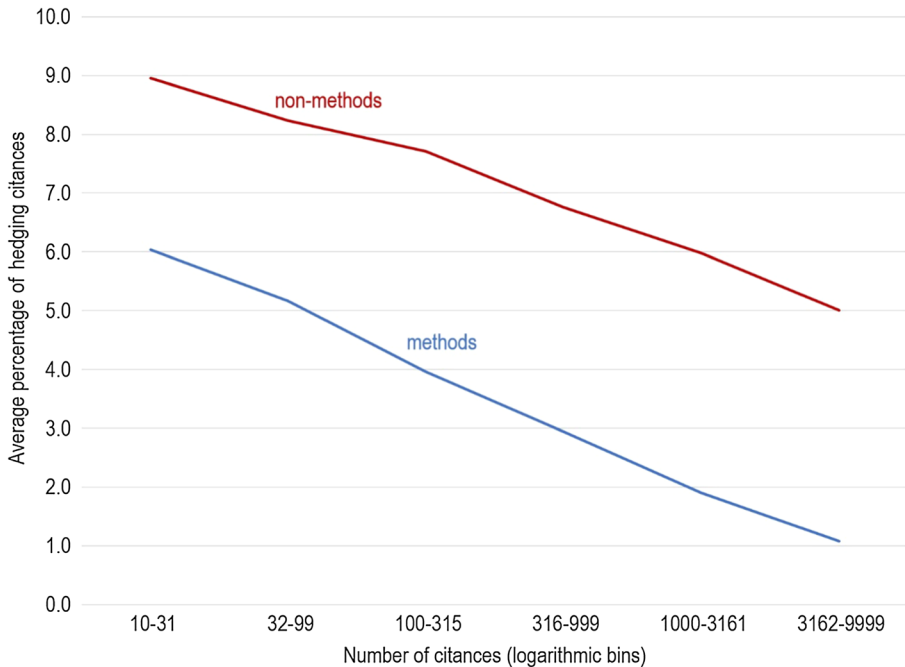
Fig. 4 Average percentage citances with hedging words as a function of number of citances (logarithmic bins) for methods and non-methods papers. The ratio of number of hedging citances to total citances is plotted for each logarithmic bin

exponential relationship to the hedging rate. Non-methods are about 4% points more highly hedged than methods, and non-methods range from almost 9% hedged at the lowest citance range to 5% at the highest, while methods range from 5.5% for the lowest range to 1% at the highest citance range.

Two types of statistical tests were used to determine if the differences in hedging rates for each of the six log10 intervals are statistically significant, either between log10 intervals, or between methods and non-methods for a given interval. If we average individual paper hedging rates for an interval we can use the non-parametric Wilcoxon–Mann–Whitney test (Teetor 2011) to determine if the distributions are shifted. Of the 15 ($n(n-1)/2$) interval comparisons for non-methods, only two were not significant at the $p < .05$ level: intervals 1 versus 2 and intervals 5 versus 6. For methods, only the interval 1 versus 2 was not significant. All six differences in averages between methods and non-methods within the same interval were significant. Using a different metric, if we measure hedging rate per interval by summing the number of hedging citances and dividing by total citances (which Fig. 4 is based on) we can test the difference of the proportions using Chi square. In all 15 plus six comparisons using Chi square, we were able to reject the null hypothesis that the proportions are the same at a p value of $< .05$.

It is not given, however, that papers with ten or fewer citances will have higher hedging rates. In fact, this trend appears to break down for papers with fewer than ten citances. Large samples of papers with citance counts of one, five and ten show aggregate hedging rates of 9.1%, 8.7% and 8.7% respectively, not differentiating between methods and
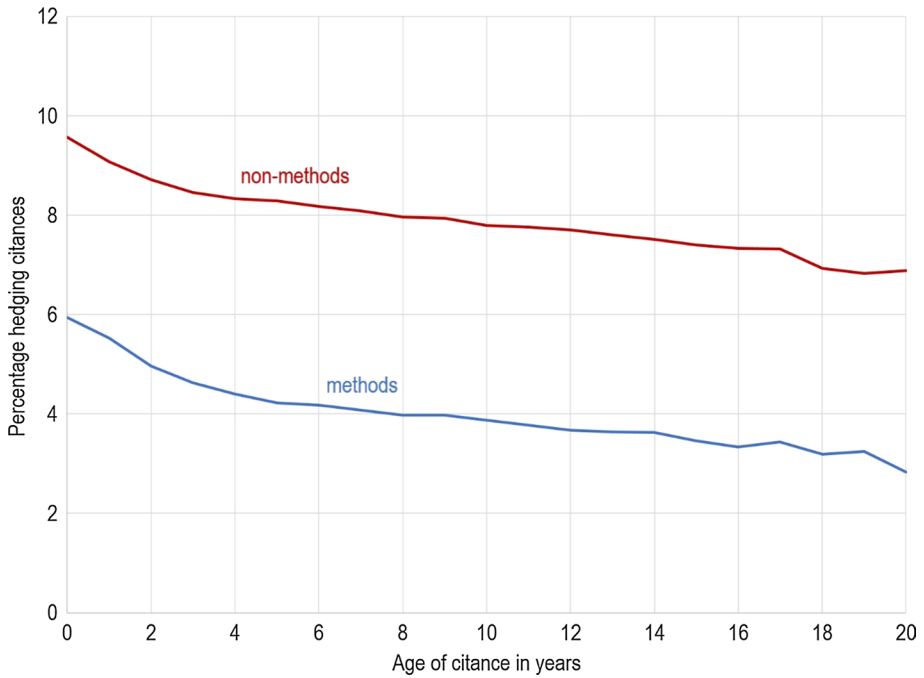
Fig. 5 Percentage of hedging citances as a function of citance age for methods and non-methods. The ratio of number of hedging citances to total citances is plotted for each citance age calculated by subtracting the publication years of the citing and cited papers

non-methods. By contrast, non-method papers in the range of 10 to 31 citances (Fig. 4) have a hedging rate of 8.9%. This suggests that there is at least a leveling-off of hedging rates for papers with fewer than 10 citances. The reason for this is not clear, but a possible explanation is self-citation, where hedging may be less likely, and will have a greater effect at these lower citation levels.

Hedging and citance age

We can also examine the time evolution of hedged citances. This analysis is based on method and non-method papers with 20 or more citances and with a publication year of 1997 or greater when the PubMed Central data set reaches critical mass. Thus, we are looking at more visible papers and how they were received immediately after publication versus later in their histories. As we might expect, early citances have a higher hedging rate than later citances for both methods and non-methods as shown in Fig. 5. Non-methods range from 9.5% hedging in the year of publication (age=0), to 6.9% at age 20 years, while methods go from 5.9% at age 0 to 2.8% at age 20. The differential between method and non-methods of about 4% was previously observed. Similar results are obtained when we select a cohort of papers published in a specific year and follow them through time. For example, non-method papers published in the year 2000 have a hedging rate of 12.3% in the year of publication which falls to 7.8% after 17 years. The most rapid decline in hedging occurs in the first 5 years after publication.

Computing hedging rate on a per word basis

One potential confounding factor in relating hedging to citance frequency is a systematic increase in citance length for lesser cited papers. A preliminary study showed that citances for less cited papers were about 6% longer than citances for highly cited papers. A longer citance is more likely to contain hedging words if such words are randomly distributed in text. Computing hedging rate on a per word basis rather than a per citance basis overcomes this potential bias because we are dividing the number of hedging words by the total number of words in the citances which are slightly longer if they refer to lesser cited papers.

We can test this hypothesis using the corpus linguistics software Wordsmith Tools (Scott 2004) by comparing highly cited and low cited samples of papers. This software computes the log likelihood statistic for words, or combinations of words called lemmas, that appear significantly more in one corpus compared to another. Two such comparisons were carried out, one for methods papers and the other for non-methods. Citances for method papers having ten citances were compared with citances for method papers with at least 1300 citances each. Similarly, for non-methods, citances for papers having 10 citances were compared with papers having at least 1270 citances. A lemma was defined containing the hedging words "may", "could", and "might". The log likelihood of the lemma for the lesser cited method sample with respect to the highly cited method sample was 3170, while the log likelihood for the lesser cited non-method sample with respect to the highly cited non-method sample was 1640. Both log likelihood values are highly significant (at the $p < .0001$ level), and thus the higher prevalence of hedging terms for lesser cited papers is confirmed using this alternative metric which takes citance length into consideration.

Discussion

One obvious limitation of this study is its restriction to the biomedical literature. We do not know if our results can be generalized to other fields and data sets. In fact, we do not know if hedging words will vary from field to field. We have also limited the analysis to a small set of hedging terms. While these common words are expected to have applicability across multiple fields, perhaps other words will assume greater importance in non-biomedical fields. In addition, to expand beyond this small set of words will require part-of-speech analysis to differentiate the hedging usage of a word like "potential", for example, from its technical usage. We have also noted a possible confounding effect of self-citations at low citance levels which will need to be examined if we are to assess the uncertainty of low cited papers.

Although we have found higher hedging rates for papers with lower citance counts, and we know that citance counts and traditional citation counts are highly correlated, we have not directly tested the hypothesis that papers with lower traditional citation counts have higher hedging rates. This could be done by aggregating citances using intervals of traditional citation counts, rather than citance count intervals as we have done. Preliminary tests indicate that the generalization will hold up using a traditional citation count definition as well. For example, papers with a traditional citation count > 500 have an average hedging rate of 3.3 compared to a rate of 8.5 for papers with 10 traditional citations across method and non-methods.

While the exploration of the hypothesis that citations are related to certainty is just getting started, it will be important to find out with a much more extensive lexicon of hedging terms just how strong this relationship is. We have observed with a small set of terms a wide range of hedging rates for individual papers from zero to 80%. Will a wider set of hedging terms steepen the curve toward lower citation rates or flatten it? Or will the difference between methods and non-methods diminish? This will depend, in part, on how much ambiguity in meaning is introduced by the additional hedging terms. It is important to emphasize that hedging does not mean that the findings of a paper are invalid, only that there are unresolved issues associated with certain of its findings. It should be possible to apply more sophisticated linguistic tools to clearly identify what these issues are for individual papers.

These linguistic tools can also be used to further differentiate the non-methods papers to determine if there are pockets of higher uncertainty. For example, we are exploring the possibility of using similar linguistic tools to identify discovery papers (Small et al. 2017) and review papers as a subset of non-method papers. We expect that discovery papers will have higher levels of uncertainty. But what does it mean if a review paper is indicating uncertainty? Unlike method papers (which refer to specific methods) and discovery papers (which refer to specific discoveries), review papers refer to a topic (a set of related papers). What does it mean that a topic is uncertain? Does the history of a topic go from uncertain to certain? Or is topic uncertainty a signal that there is an increased possibility for future discoveries and future developments in methodology in that topic (i.e. an emergent topic)?

Overall, our findings shed new light on the standard interpretations of citation counts. Up to now we have relied on indirect correlational evidence at the aggregate level that citations are related to the quality, importance, or influence of the underlying work. Our results do not invalidate these earlier interpretations because, for example, it is possible that papers of high certainty are also seen as important or of high quality. But if "importance" or "peer recognition" are just surrogates for the certainty of knowledge, then a major reorientation of these interpretations is in order. The use of citations in evaluative studies will need to take this perspective into consideration when drawing conclusions based on citation counts.

Conclusions

By reinterpreting citation counts in terms of the certainty or uncertainty of the scientific knowledge produced, we have access to scientists' perceptions of the validity of knowledge claims, that is, what contributions add to our store of knowledge and are more likely to be correct. This information can, in turn, guide scientists to findings and methods on which they can reliably build knowledge. At the same time, the relation of citations to certainty moves scientometrics and informetrics closer to epistemology and the philosophy of science. New questions arise regarding what makes one paper or finding more certain than another, and this in turn requires us to pay closer attention to the evidentiary basis of scientific findings and how science is verified. Finally, the often ignored and undervalued "method paper" assumes greater significance as a beacon of certain knowledge.

References

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.

Boyack, K. W., Van Eck, N. J., Colavizza, G., & Waltman, L. (2018). Characterizing in-text citations in scientific articles: A large-scale analysis. *Journal of Informetrics, 12*(1), 59–73.

Chen, C., & Song, M. (2018). *Representing scientific knowledge: The role of uncertainty*. London: Springer.

Cole, S. (1992). *Making science: Between nature and society*. Cambridge, MA: Harvard University Press.

Cole, J. R., & Cole, S. (1973). *Social stratification in science*. Chicago: The University of Chicago Press.

DiMarco, C., Kroon, F. W., & Mercer, R. E. (2006). Using hedges to classify citations in scientific articles. In J. Shanahan, Y. Qu, & J. Wiebe (Eds.), *Computing attitude and affect in text: Theory and applications* (Vol. 20, pp. 247–263)., The Information Retrieval Series Amsterdam: Springer. https://doi.org/10.1007/1-4020-4102-0.19.

Garfield, E. (1990). The most-cited papers of all time, SCI 1945–1988. Part 1A. The SCI top 100—Will the Lowry method ever be obliterated? *Current Contents, 7*, 3–14. February 12, 1990. [Reprinted in: *Essays of an information Scientist*, Vol. 13 (p. 45). Philadelphia, PA: ISI Press].

Gilbert, G. N. (1977). Referencing as persuasion. *Social Studies of Science, 7*(1), 113–122.

Hyland, K. (2009). *Academic discourse: English in a global context*. London: Continuum.

Latour, B., & Woolgar, S. (1979). *Laboratory life: The social construction of scientific facts*. Beverly Hills, CA: Sage Publications.

Nakov, P., Schwartz, A., & Hearst, M. (2004). *Citances: Citation sentences for semantic analysis of bioscience text*. SIGIR workshop of search and discovery on bioinformatics.

Narin, F. (1976). *Evaluative bibliometrics: The use of publication and citation analysis in the evaluation of scientific activity*. Cherry Hill, NJ: Computer Horizons Inc.

Scott, M. (2004). *WordSmith tools version 4*. Oxford: Oxford University Press.

Small, H. (1977). A co-citation model of a scientific specialty: A longitudinal study of collagen research. *Social Studies of Science, 7*, 139–166.

Small, H. (2018). Characterizing highly cited method and non-method papers using citation contexts: The role of uncertainty. *Journal of Informetrics, 12*(2), 461–480.

Small, H. & Klavans, R. (2011). Identifying scientific breakthroughs by combining co-citation analysis and citation context. In *Proceedings of the 13th international conference of the international society for scientometrics and informetrics*, Durban, South Africa.

Small, H., Tseng, H., & Patek, M. (2017). Discovering discoveries: Identifying biomedical discoveries using citation contexts. *Journal of Informetrics, 11*(1), 46–62.

Teetor, P. (2011). *R Cookbook*. Sebastopol, CA: O'Reilly Media Inc.

Van Noorden, R., Maher, R. B., & Nuzzo, R. (2014). The top 100 papers. *Nature, 514,* 550–553.

Virgo, J. A. (1977). A statistical procedure for evaluating the importance of scientific papers. *The Library Quarterly, 47,* 415–430.