# Properties of an index of citation durability of an article

Natsuo Onodera*

*Professor Emeritus, The University of Tsukuba, Kasuga, Tsukuba, Ibaraki 305-8550, Japan*

## A R T I C L E   I N F O

## A B S T R A C T

Citation Delay ($D$) introduced by Wang et al. (2015) is a measure of citation durability of articles reflecting information on the entire citation life-time. The characteristics of the measure and relationships of it to other article characteristics are examined in the six different fields using the citation data over 15 years of the articles published in 2000 in these fields. $D$ distributes normally with good approximation and is not so much dependent on the subject field as the citation count. Although articles with higher $D$ (more lately cited) tend to gain more citations in their life-time, this relationship is not linear but the mean of citations reaches a maximum at a certain value of $D$. Multiple regression analysis explaining $D$ showed that articles with a higher Price index (i.e. citing more recent references) will receive most of the citations relatively earlier and that there is a weak tendency that articles containing more figures are cited earlier and those containing more tables are cited later. A seemingly contradictory result is found that more highly cited articles tend to have higher citation durability in individual journals while high-impact journals tend to include more articles with lower citation durability in higher proportions.

© 2016 Elsevier Ltd. All rights reserved.

## 1. Introduction

The number of citations an academic paper receives is often used as a measure of the scientific impact of the paper. However, papers with the same total citation count can show different time distribution patterns of the count. Many articles are rarely cited for some time after publication, then receive a growing number of citations to arrive at a peak somewhere between two and six years after publication, before the citation count decreases, while some receive most of the citations within a year or two, others are cited constantly for a long period, and still others remain unmarked before a sudden wave of citations arrives seven or ten years afterwards.

How citation counts change over time, in other words, ageing (obsolescence) or durability of citations has been studied from various viewpoints (see Section 2 for details). However, knowledge about citation durability of articles is less accumulated compared with that about citation count itself. One of the essential reasons for it is thought to be that any quantitative index for measuring citation durability is not established yet.

The classification into early-cited (flash in the pan), delayed-cited (sleeping beauty), and normal is often used (Garfield 1980; Glänzel, Schlemmer, & Thijs, 2003; van Dalen & Henkens 2005; van Raan 2004), but the criteria of the classification differ by the authors. Some aging parameters such as the cited half-life (median of citation age distribution) do not reflect the entire pattern of life-time citations. In contrast, the Citation Delay introduced by Wang, Thijs, and Glänzel (2015) is

---

* Correspondence to: 587-3, Konakadai-cho, Inage-ku, Chiba 263-0044, Japan.
*E-mail address:* nt.onodera@y5.dion.ne.jp

the measure of citation durability reflecting the entire life-time citation information, but little has been known about its properties.

Although the systematic investigations comparing the relation of citation durability with other characteristics of articles are not so many because of a lack of the established index of citation durability, several works have reported that delayed recognized papers tend to receive more citations in the long run than early recognized ones (Aversa 1985; Costas, van Leeuwen, & van Raan, 2010; Levitt & Thelwall 2008; Levitt & Thelwall 2009; Line 1984; Wang 2013). It is, however, not clear if this conclusion applies to any subject field because the samples employed in these investigations were either limited to a small number of highly-cited papers (HCPs) or composed of papers from various fields.

The relation of citation durability with characteristics other than citedness of articles has been addressed by only a few studies. Among those, van Dalen and Henkens (2005) and Costas et al. (2010) classified papers into four classes, i.e., [I] few-cited, [II] early-cited, [III] delayed-cited and [IV] normal, and compared the characteristics of the papers within each class. Wang et al. (2015) investigated the influence of interdisciplinarity of an article and some article features on the Citation Delay mentioned above. However, van Dalen and Henkens (2005) only made comparison of each class II, III and IV with the class I, and did not explicitly show any difference among the former three. In Costas et al. (2010) and Wang et al. (2015), the relationships between citation durability and various properties of articles were reported rather briefly because their studies mainly focused on other issues (development of general methodology for the classification of research publications according to the citation durability in Costas et al., and examination of the relation between citation durability and interdisciplinarity of articles in Wang et al.).

Citation Delay introduced by Wang et al. (2015) is considered to be the most appropriate measure for citation durability among those that have ever been proposed. Using this measure as an index of citation durability of articles, the present work sets up its purpose as follows:

(1) Elucidating the characteristics of the distribution of Citation Delay;
(2) Examining relationships between this durability index and the citation count of articles in different subject fields to reveal whether or not there is any difference in the long-range citation counts between early-cited and delayed-cited articles; and
(3) Examining relationships between the durability index and other characteristics of articles in different fields to reveal tendencies common to these fields.

In Section 2 I briefly review previous studies related to the present work. Section 3 explains the durability index, the data and method used in the subsequent analysis. Section 4 and Section 5 describe the results and discussion, respectively, and Section 6 gives main conclusions obtained from this research.

## 2. Literature review

### 2.1. Developing mathematical models for aging of papers

Studies on the obsolescence function that describes temporal changes of the citation count has been made for a long time. The most simplistic approach is to fit it to an exponentially decaying curve, but it is not adequate even qualitatively because the citation count is generally recognized to reach its peak a few years after publication.

Avramescu (1979) proposed two types of ageing function $c(t)$ describing the citation count after a lapse of time $t$, which can fairly approximate the citation history of numerous articles by adjusting three parameters in those functions. Egghe and Ravichandra Rao (1992) examined the ageing factor $a(t) = c(t+1)/c(t)$ and proposed a log-normal model for $c(t)$ based on the empirical observations that in many cases $a(t)$ has a minimum at a certain $t$. Burrell (2003) analyzed citation age distribution (yearly change of citations received by articles) based on the failure rate function in the reliability theory, and supported the conclusion by Egghe and Ravichandra Rao (1992).

Glänzel and Schoepflin (1995) used a stochastic model for the process to acquire citations from a set of articles (e.g., those published in a certain journal in a certain year) and defined indices for the speed of early reception and for later ageing. Based on applications of these indices to actual cases, they showed that the ageing patterns depend on discipline rather than journal, and that slow ageing does not necessarily mean slow reception.

Della Briotta Parolo et al. (2015) investigated the change of citation decay with publication year for articles published during 1960–1990 and found that the more recently the articles were published, the faster they reach the citation peak and also the shorter their citation half-life becomes. In addition, they showed that the citation trajectories after the peak year fit better to exponential decay than power law decay.

### 2.2. Comparison of citation durability between highly cited papers (HCPs) and other papers

Early studies by Line (1984) demonstrated that highly cited papers (HCPs) have longer citation durability. This findings raised interest in the citation age distribution of HCPs.

Aversa (1985) and Cano and Lind (1991) characterized the citation ageing pattern for a long period of HCPs they selected, i.e., 400 HCPs published in 1972 in case of Aversa and ten HCPs from two disciplines chosen by Garfield as "citation classics"

(Garfield, 1984a, 1984b) in case of Cano and Lind, and both showed that the papers could be classified into two groups with short and long citation periods. Aversa also indicated that the papers in the latter group show higher total citation counts than the former. In addition, Aksnes (2003) found that the citation age of the HCPs is somewhat longer than the other papers from an analysis of articles published by Norwegian authors during 1981–89.

Levitt and Thelwall (2008) addressed the issue on the temporal changes of citation counts for HCPs in a more systematic manner. Based on ageing patterns for 100 HCPs published in 1974, they claimed that the patterns were highly diverse and far more complicated than dividing them into two groups as shown by Aversa (1985) and Cano and Lind (1991). They further analyzed the citation age distribution during 36 years for 36 HCPs each from six disciplines published in 1969–1971, and concluded that: (a) the ratio of the early citation count (in the initial six years) to the total citation count varied remarkably among the articles regardless of the discipline; (b) articles with higher total citation counts tended generally to show fewer early citations, but exact correlation was dependent on the discipline; and (c) the citation durability of HCPs was higher than non-HCPs for all the disciplines. Levitt and Thelwall (2009) found also for HCPs in the field of information and library science that there is a moderate correlation indicating a delayed citation tendency for the articles ranked as more highly cited.

### 2.3. The change in the citation ranking of articles when changing the citation window length

Abramo, Cicero, and D'Angelo (2011) estimated the error in measurement that occurs with decreasing the citation window length for eight different disciplines, based on the citation data till the end of 2008 of papers published in 2001 by Italian authors. For the 3-years citation window including the publication year, for example, the cumulative citation counts amounted to 65–88% (depending on the discipline) of the final (9-years citation window) counts and Spearman's rank correlation coefficients for the citation rankings between the 3-years and 9-years citation windows were 0.79–0.96.

When studying the correlation between short and long citation impacts of papers, Wang (2013) introduced "Citation Speed" as an indicator to measure how fast an article acquires its citations. Let the publication year of an article be 0 and the cumulative citation count from year 0 to year $t$ be $C(t)$, then the Citation Speed of the article at year $T$ is given by:

$$\text{Citation Speed} = \frac{1}{T} \sum_{t=0}^{T-1} C(t)/C(T)$$

The Citation Speed lies between 0 and 1 and the earlier an article is cited, the more closely it gets near to 1. From the observation of the citation data during 31 years of articles published in 1980, Wang showed that the distribution of the Citation Speed of articles is heterogeneous not only across subject fields but also within a subject field or even within a journal, meaning considerable difference in the citation rankings between the short-term and the long-term citation windows. He also indicated that the more highly cited articles in its life-time tend to receive more delayed citations.

Both Abramo, Cicero et al. (2011) and Wang (2013) indicated the problem of evaluating the impact of articles using citation data in a short period after publication and claimed that the citation window in research assessment should be carefully chosen considering the trade-off between accuracy and timeliness in measurement.

Levitt and Thelwall (2008) reported that the correlation between the citation rank within HCPs in the first six years and that in the entire period (36 years) was insignificant in five out of six disciplines they surveyed (see Subsection 2.2).

### 2.4. Classification of papers by their citation durability and indicators for measuring citation durability

Some researchers have attempted to classify articles with respect to the citation durability. Particularly, articles that attract attention after a prolonged period without citations, called "delayed recognition papers" (Garfield 1980; Glänzel et al., 2003) or "sleeping beauties" (van Raan 2004), are a focus of interest. Conversely, papers frequently cited immediately after publication and then forgotten are called "flashes in the pan" (van Dalen & Henkens 2005). Articles that have been regarded as "flashes in the pan" in the past suddenly starting to be cited at some point in time were recently named "all-elements sleeping beauties" (Li, 2014; Li & Ye 2012).

Glänzel et al. (2003) examined the citation history for 21 years for papers published in 1980 and identified 60 "delayed recognition papers". On the other hand, van Raan (2004) defined the "sleeping beauties" according to (a) the maximum citations per year during sleeping, (b) the sleeping period, and (c) the minimum citations per year after awakening, and identified about 360 papers meeting this definition from those published in and after 1980. Although the delayed recognition papers (or sleeping beauties) identified by these two studies look rather rare, Burrell (2005) demonstrated that they appear with a frequency higher than that assuming statistically random events based on calculations using his citation ageing model, and suggested that there must be some reasons for delayed recognition.

Li, Shi, Zhao, and Ye (2014) addressed the citation age distribution during the sleeping period, which they called "heartbeat spectrum", of sleeping papers and revealed that papers with "late heartbeats" have much higher awakening probability (more likely to become sleeping beauties) than those with "early heartbeats."

Lachance and Larivière (2014) extracted "sleepers" from papers published during 1963–1975, as papers which received no citation for ten years or more since publication but obtain some citations hence, and compared citation pattern of the sleepers (about 5% of the all papers) with that of the non-sleeper papers of the same period (control group). Although the citation age distribution of sleepers after awakening showed a gradual decline in general similar to the control group, but

the highly-cited sleepers (i.e., sleeping beauties) showed no decline but even increase, differing from patterns of the control group. It is noted that Lachance and Larivière added that these results are difficult to generalize since the size of the highly-cited sleepers is very small and also citations received by these papers are much less than those by highly-cited papers in the control group.

Li and Shi (2016) claimed that the criteria for sleeping beauties by van Raan (2004) are not applicable at present and proposed two new criteria based on much longer citation history. From articles of Nobel Prize laureates with at least 50 years citation history they extracted 25 sleeping beauties according to their criteria, finding that only 10 of them met van Raan's criteria.

Baumgartner and Leydesdorff (2014) proposed a method of dividing a set of papers into subsets according to the distribution pattern of citation age using the group-based trajectory modeling (GBTM), a non-parametric statistical technique. Application of this method to the 16-years' citation history of several sets of papers published in 1996 yielded three to seven groups which are characterized by not only the citation frequency but the citation durability. However, the authors admit the difficulty of application of the method to large sets of papers from a number of journals because the number of groups obtained depends on subjective judgment and because the model requires simplification by eliminating outliners and defining initial values.

Wang et al. (2015) defined "Citation Delay" by subtraction of the Citation Speed (see Subsection 2.3) from one and analyzed how this measure calculated using long-term (13 years) citation data of articles published in 2001 is influenced by the interdisciplinarity of the articles and some other article features. Besides Citation Delay, no quantitative indicator of citation durability reflecting information on the entire citation age distribution has been proposed. Although the index $G_s$ introduced by Li et al. (2014) is very similar to Citation Delay, they used it only for measuring inequality of the "heartbeat spectrum" (see above), and not as a general indicator of citation durability.

On the other hand, Ke et al. (2015) introduced the "Beauty Coefficient" indicator which measures simultaneously the depth of sleeping and intensity of awaking. This indicator measures the gap between the number of citations at the year of the citation peak and the yearly citations until the peak year. The distributions of this indicator for citation data of two data sets for a long period were subject to the power law. They concluded, from these distributions, there are no clear demarcation values to separate "sleeping beauties" from "normal" papers.

### 2.5. Relationships between citation durability and other characteristics of articles

The systematic comparison of the relation of citation durability with other characteristics of articles has been investigated by only a few studies: van Dalen and Henkens (2005), Costas et al. (2010), and Wang et al. (2015). These studies are already outlined in Section 1, but will be discussed later in more detail (see Subsection 5.3) in relation to our present work.

## 3. Data and method

### 3.1. The index of citation durability used

The Citation Delay (referred to symbol $D$ hereinafter), which was introduced by Wang et al. (2015) and is used as the index of citation durability in this work, was described briefly in Section 1. It is defined again as follows:

Let the publication year of an article be $j = 0$, and the citation count within each of the years $j = 0, 1, 2, \ldots T$ be $c(j)$, where $T$ is the last year in which the citations are observed. Then the cumulative citation count $C(t)$ from the publication year to $j = t$ is given by

$$C(t) = \sum_{j=0}^{t} c(j) \tag{1}$$

The relative cumulative citation count $x(t)$ up to $j = t$ is

$$x(t) = C(t)/C(T) \tag{2}$$

where $C(T)$ is the total number of citations received between $j = 0$ and $j = T$. $D$ of the article at $j = T$ is defined as:

$$D = 1 - \frac{1}{T} \sum_{t=o}^{T-1} x(t) \tag{3}$$

This definition is illustrated in Fig. 1, where the horizontal axis represents $t$ ($0 \leqq t \leqq T - 1$) and the vertical axis $x(t)$. The area of the shaded part relative to the total area of the rectangle corresponds to $D$.

If a paper receives citations in the year of publication ($t = 0$) only and none subsequently, $x(t)$ reaches 1 at $t = 0$ and remains there, which means that the area of the shaded part ($D$) is zero. If, conversely, the citation count remains zero from $t = 0$ to $t = T - 1$ and assumes a certain value at $t = T$, the total area of the rectangle of Fig. 1 is shaded, and therefore $D = 1$. Therefore, $D$ of an article lies between 0 and 1; earlier citations lead to smaller areas of the shaded part (i.e. smaller $D$) and later citations
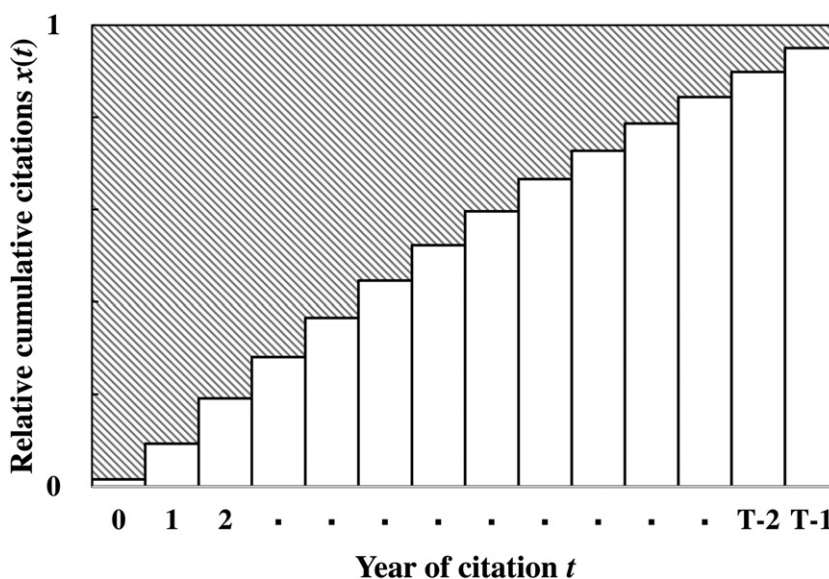
**Fig. 1.** Illustration of the Citation Delay $D$.    When the area of the rectangle is 1, the area of shaded part is $D$.

result in larger $D$. Thus, $D$ can be considered as an index of citation durability reflecting information on the entire citation age distribution.

$D$ depends on the relative shape of the graph shown in Fig. 1 but not on the total number of citations. It should be noted that $D$ is not defined for $C(T) = 0$ and little meaningful for very low $C(T)$. The present work is principally concerned with papers with $C(T)$ of 5 or more for 15 years.

### 3.2. Data used for analysis

As described in Section 1, this paper analyzes the several characteristics of $D$ in different fields and examines whether there are general tendencies beyond these fields. According to this research purpose, I divided all the Subject Categories used in the Web of Science (WoS) into eleven broad areas and identified six broad areas containing more journals which are Physics, Chemistry, Engineering, Biology, Basic medicine, and Clinical medicine. From each of the six, I selected one Subject Category by considering that: (a) it can be regarded as a representative of the broad area, (b) it has more than 5000 papers annually, and (c) it has the weighted mean impact factor of the category near the mean of the broad area. The selected six Subject Categories are as follows:

– Physics: Condensed-matter physics (CondMat)
– Chemistry: Inorganic and nuclear chemistry (Inorg)
– Engineering: Electric and electronic engineering (Elec)
– Biology: Biochemistry and molecular biology (Biochem)
– Basic medicine: Physiology (Physiol)
– Clinical medicine: Gastroenterology and hepatology (Gastro)

These Subject Categories are called "fields" and the abbreviations shown in parentheses will be used hereinafter.
Four journals are chosen from each field (see Table 1). The selection is made taking the following into account:

(a) Journals classified only in the relevant categories (nearly 50% of the journals included in WoS are classified in two categories or more).
(b) Journals using English only
(c) Journals published in different countries included in each field
(d) Journals with relatively high and low impact factors included

All of the normal articles (assigned the document type "article" in WoS) published in 2000 in the 24 journals chosen above were extracted for analysis, excluding those simultaneously classified as "proceedings paper" and those without author's name. The numbers of the papers from each journal are also shown in Table 1.

The reason that the journals and papers for analyses are selected under the conditions and restrictions mentioned above is to minimize the possible influences of factors other than those investigated in the analyses. Even within the same WoS

**Table 1**
Selected subject fields and journals.

| Subject field | Journal title[a] | Publishing country | Sampled articles | Journal code[b] |
|---|---|---|---|---|
| Condensed Matter Physics (CondMat) | European Physical Journal B | DEU | 538 | |
| | Journal of Physics – Condesed Matter | GBR | 813 | CondMat-1 |
| | Physica B | NLD | 148 | |
| | Physical Review B | USA | 4738 | CondMat-2 |
| Inorganic and Nuclear Chemistry (Inorg) | Inorganic Chemistry | USA | 931 | Inorg-1 |
| | Inorganica Chimica Acta | CHE | 546 | |
| | Journal of the Chemical Society – Dalton Transactions | GBR | 682 | Inorg-2 |
| | Transition Metal Chemistry | NLD | 139 | |
| Electric and Electronic Engineering (Elec) | IEE Proceedings – Circuits, Devices and Systems | GBR | 52 | |
| | IEEE Transactions on Circuits and Systems I | USA | 218 | |
| | IEEE Transactions on Microwave Theory and Techniques | USA | 295 | Elec-1 |
| | Signal Processing | NLD | 178 | Elec-2 |
| Biochemistry and Molecular Biology (Biochem) | European Journal of Biochemistry | GBR | 788 | Biochem-1 |
| | Journal of Biochemistry (Tokyo) | JPN | 275 | |
| | Journal of Biological Chemistry | USA | 5504 | Biochem-2 |
| | Journal of Molecular Biology | USA | 875 | |
| Physiology (Physiol) | Journal of General Physiology | USA | 110 | Physiol-1 |
| | Japanese Journal of Physiology | JPN | 72 | |
| | Journal of Physiology – London | GBR | 472 | Physiol-2 |
| | Pflugers Archive European Journal of Physiology | DEU | 238 | |
| Gastroenterology (Gastro) | American Journal of Gastroenterology | USA | 430 | Gastro-1 |
| | Gastroenterology | USA | 259 | |
| | Gut | GBR | 277 | Gastro-2 |
| | Journal of Gastroenterology | JPN | 124 | |

[a] The journal titles at the time of 2000, although some were changed after that.
[b] The 12 journals coded here are analyzed at journal level (see the Subsection 3.2). These codes are used in the following text.

Subject Category, citation frequency and other properties of an article have different tendencies depending on subdomain, document type, and language. This study aims to investigate similarity or difference among the six fields by eliminating the effects of those factors as much as possible. Besides, journals with relatively low impact factors are also included to make samples represent all articles in the field.

The citation data were downloaded from WoS on March 31st, 2015 and the citation counts were recorded for every year from 2000 ($t = 0$) to 2014 ($t = T = 14$).

Analyses were conducted for each field and additionally for each of the 12 journals whose codes are shown in the fifth column of Table 1.

### 3.3. Calculation of D

Out of 18,702 articles contained in the 24 journals, 331 were never cited in the whole period of 2000–2014, i.e. give $C(T) = 0$. $D$ was calculated using the Eqs. (1) to (3) for the remaining 18,371 articles. It should be noted that 1556 articles among them have $C(T)$ as low as 1–4 and consequently are not useful for discussing the citation durability. Analysis was therefore chiefly concerned with the remaining 16,815 articles with $C(T) \geqq 5$.

Little is known how $D$'s values distribute in a given set of articles. (Wang et al. (2015) claimed "Citation Delay is roughly normally distributed" but did not show any data for supporting it.) In this study, the distribution of $D$ was examined and compared to a normal distribution for each of the six fields and the 12 journals whose codes are shown in Table 1.

### 3.4. Correlation of citation durability with total number of citations

The correlation of $D$ with $C(T)$ was studied for articles with $C(T) \geqq 5$ in the six fields and 12 journals. Since the distribution of $C(T)$ is highly skewed, the logarithmic transform $\log(C(T) + 1)$ was used for calculation of Pearson's product-moment correlation coefficients with $D$.

### 3.5. Multiple regression analysis for explanation of D from characteristics of articles

In order to find characteristics of articles significantly correlated with $D$, multiple regression analysis was performed with $D$ as the response variable and characteristics shown in Table 2 as explanatory variables. As a preliminary step, Pearson's product-moment correlation coefficients between $D$ and the explanatory variables were calculated.

The explanatory variables used here include the total number of citations $C(T)$ and the impact factor ($IF$) of the journal in which the article appeared, in addition to those the author used in the study of factors influencing the citation counts (Onodera & Yoshikane 2015). The first two were added because of their significant correlation with $D$ (see Subsection 4.3 and Subsection 4.4).

**Table 2**
Explanatory variables for the multiple regression analysis predicting *D*.

| Variable | Definition | Log transform | Data acquisition |
|---|---|---|---|
| Authors | Number of authors of the article | No transform | Count entries in the AU field of the WoSCC. |
| Insts | Number of institutions with which the authors are affiliated | No transform | Count entries in the C1 field of the WoSCC. |
| Countries | Number of countries where the institutions are located | No transform | Count from country names at the end of the C1 entries of the WoSCC. |
| Refs | Number of references cited in the article | No transform | Obtain from the WoSCC NR field. |
| Price | Price index (percentage of the refffferences whose publication year is within 5 years before the publication year of the article) | No transform | Calculate based on the reference list in the WoSCC CR field. |
| Length | Number of normalized pages of the article | No transform | Count directly from the original document and normalize the number of pages as 6400 characters per page. |
| Figures | Number of figures in the article | No transform | Count directly from the original document (include figures, charts, diagrams, pictures, and also tables presented as figures). |
| Tables | Number of tables in the article | No transform | Count directly from the original document. |
| Eqs | Number of numbered equations in the article | Log of the value plus 1 | Count directly from the original document. |
| Age | Active years (elapsed years from the year of the first article publication to the year 2000) of the first author | Log of the value plus 1 | Analyze the results of the WoSCC author name search during 1970–2000.[a] |
| RatePubl | Number of articles published per annum by the first author during his (her) active years | Log of the value plus 0.1 | Analyze the results of the WoSCC author name search during 1970–2000.[a] |
| MedCite | Median of the number of citations received per annum by each published article | Log of the value plus 0.01 | Calculate based on the citation data of the articles obtained from the analysis of the author name search abovementioned. |
| C(T) | Number of citations received by the article till the end of 2014 | Log of the value plus 1 | See the Subsection 3.2. |
| IF | Impact Factor of the journal in which the article was published (average of the values of 2001 and 2002) | No transform | Obtain from the Journal Citation Reports. |

[a] See Onodera et al. (2011) about details for the analysis.

Logarithmic transforms were used for variables *C*(*T*), *Eqs*, *Age*, *RatePubl* and *MedCite* because they showed a skewness of 2 or higher, or a ratio of the mean to the median of 1.5 or higher, in more than half of the subject fields. The skewness is obtained by the SKEW function of Excel, which calculates the unbiased estimator of the population skewness:

$$\frac{n}{(n-1)(n-2)}\sum_{i=1}^{n}\left(\frac{x_i-\bar{x}}{s}\right)^3$$

Since each of these variables may have a value of zero for certain articles, 1 was added to *C*(*T*), *Eqs* and *Age*; 0.1 to *RatePubl*; and 0.01 to *MedCite*, before transformation (see the third column of Table 2). The values added were chosen considering the distribution of each variable.

The explanatory variables are acquired as shown in the fourth column of Table 2. As seen there, many of the variables cannot be obtained only from data downloaded from WoS. Specifically *Age*, *RatePubl*, and *MedCite* are known only by counting all the publications of the same author and the citations they have received (Onodera et al., 2011). Since it is unpractical to do so for all the 16,815 papers with $C(T) \geqq 5$, 60 articles or less each randomly selected from 24 journals were used in the multiple regression analysis. The number of sampled articles is shown in Table 3.

SPSS Statistics Base 17.0 was used for multiple regression analysis. Statistically significant explanatory variables were selected in a stepwise process.

## 4. Results

### 4.1. Distribution of total number of citations C(T)

The distribution of *C*(*T*) needs to be described before discussing the properties of *D*.

Table 4 shows basic statistics of the total number of citations *C*(*T*) along with the number of articles with $C(T) \geqq 1$ and $C(T) \geqq 5$ for the six subject fields as well as 12 journals (2 from each field). It is well known that the distribution of citation counts is highly skewed even within a field or a journal; the distribution of *C*(*T*) has the same character.

Table 5 shows the distribution statistics of the logarithmic transform log[*C*(*T*) + 1] (1 is added to consider the existence of articles with *C*(*T*) = 0). While the mean and median of *C*(*T*) are very different (mean is 1.5–2 times greater than median), those of log[*C*(*T*) + 1] are almost identical. The skewness of log[*C*(*T*) + 1] is far lower than that of *C*(*T*). These facts suggest a nearly symmetrical distribution of log[*C*(*T*) + 1] around the mean. Fig. 2 shows Q-Q plots to examine whether the distribution of

**Table 3**
Number of articles sampled for the multiple regression analysis.

| Field | Journal | Sampled articles | |
|---|---|---|---|
| CondMat | Eur Phys J B | 49 | 177 |
| | J Phys-Condens Matter | 42 | |
| | Physica B | 39 | |
| | Phys Rev B | 47 | |
| Inorg | Inorg Chem | 50 | 192 |
| | Inorg Chim Acta | 53 | |
| | J Chem Soc-Dalton Trans | 49 | |
| | Transit Met Chem | 40 | |
| Elec | IEE Proc-Circuit Device Syst | 25 | 145 |
| | IEEE Trans Circuits SystI-Fundam Theor Appl | 50 | |
| | IEEE Trans Microw Theory Tech | 37 | |
| | Signal Process | 33 | |
| Biochem | Eur J Biochem | 57 | 231 |
| | J Biochem (Tokyo) | 54 | |
| | J Biol Chem | 60 | |
| | J Mol Biol | 60 | |
| Physiol | J Gen Physiol | 34 | 199 |
| | Jpn J Physiol | 60 | |
| | J Physiol-London | 55 | |
| | Pflugers Arch | 50 | |
| Gastro | Am J Gastroenterol | 47 | 199 |
| | Gastroenterology | 56 | |
| | Gut | 52 | |
| | J Gastroenterol | 44 | |

**Table 4**
Statistics on the total number of citations $C(T)$.

(a) For 6 subject fields

| Field | #Articles | Mean | S.D. | Median | Skewness[a] | #Articles with $C(T) \geq 1$ | #Articles with $C(T) \geq 5$ |
|---|---|---|---|---|---|---|---|
| CondMat | 6237 | 30.0 | 72.8 | 15 | 33.1 | 6028 | 5097 |
| Inorg | 2298 | 28.9 | 33.5 | 19 | 4.8 | 2278 | 2061 |
| Elec | 743 | 19.6 | 36.5 | 9 | 5.9 | 698 | 517 |
| Biochem | 7442 | 63.5 | 78.1 | 44 | 12.5 | 7432 | 7327 |
| Physiol | 892 | 38.9 | 51.9 | 26 | 7.4 | 876 | 808 |
| Gastro | 1090 | 60.9 | 81.9 | 38 | 4.0 | 1059 | 1005 |

(b) For 12 journals

| Journal | #Articles | Mean | S.D. | Median | Skewness[a] | #Articles with $C(T) \geq 1$ | #Articles with $C(T) \geq 5$ |
|---|---|---|---|---|---|---|---|
| CondMat-1 | 813 | 15.5 | 22.5 | 9 | 4.0 | 768 | 570 |
| CondMat-2 | 4738 | 34.2 | 81.9 | 18 | 30.4 | 4610 | 4046 |
| Inorg-1 | 931 | 36.5 | 35.1 | 26 | 2.5 | 929 | 882 |
| Inorg-2 | 682 | 30.2 | 37.8 | 20 | 6.8 | 680 | 626 |
| Elec-1 | 295 | 21.9 | 35.0 | 12 | 4.6 | 284 | 228 |
| Elec-2 | 178 | 9.9 | 15.0 | 5 | 3.3 | 161 | 97 |
| Biochem-1 | 788 | 32.4 | 45.5 | 23 | 12.5 | 785 | 754 |
| Biochem-2 | 5504 | 70.5 | 67.4 | 51 | 3.2 | 5501 | 5469 |
| Physiol-1 | 110 | 50.9 | 37.2 | 44.5 | 1.1 | 109 | 107 |
| Physiol-2 | 472 | 50.3 | 64.1 | 34 | 6.9 | 470 | 459 |
| Gastro-1 | 430 | 43.3 | 46.6 | 29.5 | 2.9 | 408 | 385 |
| Gastro-2 | 277 | 65.4 | 78.0 | 44 | 5.1 | 275 | 269 |

[a] Obtained by the SKEW function of Excel (see the Subsection 3.5).

$\log[C(T)+1]$ is close to a normal one. The distribution curves for the six fields seem to be nearly linear in the range of normal theoretical quantile $[-2, 2]$ (roughly corresponding to the percentile range [2.5, 97.5]), except that slight curving is seen for 'Gastro' field. Therefore, statistical analyses on the supposition of the normality (Pearson's product-moment correlation analysis and linear multiple regression analysis) are applied to $\log[C(T)+1]$ in the following subsections.

### 4.2. Distribution of D

#### 4.2.1. Overview of distribution

Fig. 3 shows the distributions of $D$ calculated from the Eq. (3) for articles with $C(T) \geq 5$ in the six fields. These distributions are approximately symmetrical around $D = 0.4$–$0.5$. More specifically, for the six fields and 12 journals (2 from each field), Table 6 and Table 7 show the following statistics of $D$ of articles with $C(T) \geq 1$ and $C(T) \geq 5$ respectively,:

**Table 5**
Statistics on log[$C(T)+1$].

(a) For 6 subject fields

| Field | #Articles | Mean | S.D. | Median | Skewness |
|---|---|---|---|---|---|
| CondMat | 6237 | 1.20 | 0.51 | 1.20 | −0.1 |
| Inorg | 2298 | 1.29 | 0.42 | 1.30 | −0.3 |
| Elec | 743 | 1.00 | 0.51 | 1.00 | 0.1 |
| Biochem | 7442 | 1.64 | 0.39 | 1.65 | −0.3 |
| Physiol | 892 | 1.38 | 0.46 | 1.43 | −0.5 |
| Gastro | 1090 | 1.52 | 0.54 | 1.59 | −0.6 |

(b) For 12 journals

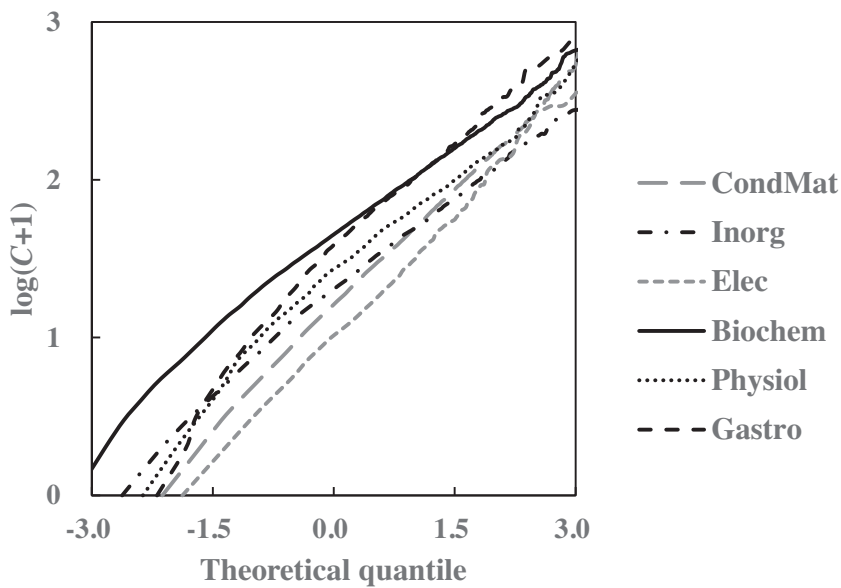| Journal | #Articles | Mean | S.D. | Median | Skewness |
|---|---|---|---|---|---|
| CondMat-1 | 813 | 1.0 | 0.5 | 1 | 0.0 |
| CondMat-2 | 4738 | 1.3 | 0.5 | 1 | −0.2 |
| Inorg-1 | 931 | 1.4 | 0.4 | 1 | −0.3 |
| Inorg-2 | 682 | 1.3 | 0.4 | 1 | −0.2 |
| Elec-1 | 295 | 1.1 | 0.5 | 1 | 0.0 |
| Elec-2 | 178 | 0.8 | 0.5 | 1 | 0.2 |
| Biochem-1 | 788 | 1.4 | 0.4 | 1 | −0.2 |
| Biochem-2 | 5504 | 1.7 | 0.4 | 2 | −0.2 |
| Physiol-1 | 110 | 1.6 | 0.4 | 1.7 | −1.0 |
| Physiol-2 | 472 | 1.5 | 0.4 | 2 | −0.3 |
| Gastro-1 | 430 | 1.4 | 0.5 | 1.5 | −0.9 |
| Gastro-2 | 277 | 1.6 | 0.4 | 2 | −0.6 |



Fig. 2. Q-Q plots for log[$C(T)+1$] of the articles with $C(T) \geqq 1$.

– Mean
– Standard deviation (SD)
– First, second, and third quartiles (P25, P50, P75)
– Quartile differences (P50 − P25, P75 − P50)

The tables indicate that:

(1) The means and medians (P50s) are slightly greater for the $C(T) \geqq 5$ group than for the $C(T) \geqq 1$ group in the fields of 'CondMat', 'Inorg' and 'Elec'. Practically no such difference is observed in the fields of 'Biochem', 'Physiol' and 'Gastro'.
(2) The standard deviations are clearly lower for the $C(T) \geqq 5$ group than for the $C(T) \geqq 1$ group (by 2–28%, depending on the subject field and journal). Similarly, the quartile differences (P50 − P25 and P75 − P50) are smaller for the $C(T) \geqq 5$ group
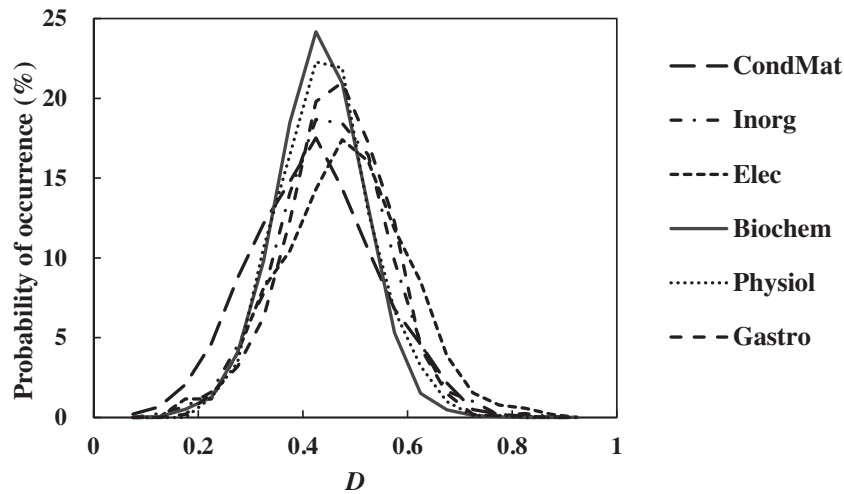
**Fig. 3.** Distribution of $D$ of articles with $C(T) \geqq 5$.

**Table 6**
Statistics on $D$ of the articles with $C(T) \geqq 1$.

(a) For 6 subject fields

| Field | $n$ | Mean | S.D. | P25 | P50 | P75 | P50−P25 | P75−P50 |
|---|---|---|---|---|---|---|---|---|
| CondMat | 6028 | 0.411 | 0.139 | 0.321 | 0.411 | 0.499 | 0.090 | 0.088 |
| Inorg | 2278 | 0.450 | 0.119 | 0.380 | 0.452 | 0.523 | 0.073 | 0.070 |
| Elec | 698 | 0.462 | 0.157 | 0.359 | 0.473 | 0.555 | 0.114 | 0.081 |
| Biochem | 7432 | 0.431 | 0.087 | 0.376 | 0.434 | 0.488 | 0.058 | 0.054 |
| Physiol | 876 | 0.441 | 0.102 | 0.382 | 0.443 | 0.500 | 0.061 | 0.057 |
| Gastro | 1059 | 0.462 | 0.106 | 0.400 | 0.465 | 0.531 | 0.065 | 0.066 |

(b) For 12 journals

| Journal | $n$ | Mean | S.D. | P25 | P50 | P75 | P50−P25 | P75−P50 |
|---|---|---|---|---|---|---|---|---|
| CondMat-1 | 768 | 0.422 | 0.157 | 0.324 | 0.429 | 0.525 | 0.104 | 0.097 |
| CondMat-2 | 4610 | 0.408 | 0.131 | 0.321 | 0.409 | 0.490 | 0.088 | 0.081 |
| Inorg-1 | 929 | 0.453 | 0.106 | 0.390 | 0.455 | 0.524 | 0.065 | 0.069 |
| Inorg-2 | 680 | 0.434 | 0.112 | 0.367 | 0.435 | 0.503 | 0.068 | 0.068 |
| Elec-1 | 284 | 0.449 | 0.143 | 0.361 | 0.462 | 0.536 | 0.101 | 0.074 |
| Elec-2 | 161 | 0.459 | 0.191 | 0.343 | 0.464 | 0.585 | 0.121 | 0.121 |
| Biochem-1 | 785 | 0.443 | 0.105 | 0.375 | 0.450 | 0.510 | 0.075 | 0.061 |
| Biochem-2 | 5501 | 0.430 | 0.083 | 0.377 | 0.433 | 0.484 | 0.056 | 0.051 |
| Physiol-1 | 109 | 0.432 | 0.075 | 0.389 | 0.423 | 0.478 | 0.034 | 0.054 |
| Physiol-2 | 470 | 0.444 | 0.093 | 0.387 | 0.446 | 0.499 | 0.059 | 0.053 |
| Gastro-1 | 408 | 0.468 | 0.112 | 0.399 | 0.476 | 0.545 | 0.077 | 0.069 |
| Gastro-2 | 275 | 0.463 | 0.102 | 0.404 | 0.472 | 0.533 | 0.068 | 0.061 |

in most fields and journals. In other words, the distribution of $D$ is narrower if only articles with greater total number of citations are considered.

(3) The means and medians of $D$ are considerably close, particularly for the $C(T) \geqq 5$ group, in any field or journal. Also the difference between P50 and P25 and that between P75 and P50 are close to each other. The visual impression of symmetrical distribution given by Fig. 3 is thus supported numerically.

### 4.2.2. Examination of normality of distribution

The ratios of P50 − P25 or P75 − P50 to the standard deviation (SD) of the distribution of $D$, as shown in Table 6 and Table 7, are in a range 0.46–0.73, which is close to the theoretical ratio of 0.674 for the normal distribution. Q-Q plots shown in Fig. 4 comparing the distributions of $D$ of articles with $C(T) \geqq 5$ to normal ones are almost linear for all the six fields. Q-Q plots for articles with $C(T) \geqq 1$ (not shown) are also nearly linear, but the slope gradually becomes steeper in the region $D > 0.6$.

It is thus concluded that $D$ follows the normal distribution fairly closely, particularly for articles with citation counts greater than a certain level, which makes $D$ a favorable quantity for statistical treatment.

**Table 7**
Statistics on $D$ of the articles with $C(T) \geqq 5$.

(a) For 6 subject fields

| Field | $n$ | Mean | S.D. | P25 | P50 | P75 | P50−P25 | P75−P50 |
|---|---|---|---|---|---|---|---|---|
| CondMat | 5097 | 0.420 | 0.119 | 0.338 | 0.419 | 0.498 | 0.081 | 0.079 |
| Inorg | 2061 | 0.456 | 0.105 | 0.388 | 0.457 | 0.522 | 0.069 | 0.065 |
| Elec | 517 | 0.480 | 0.119 | 0.399 | 0.484 | 0.554 | 0.085 | 0.069 |
| Biochem | 7327 | 0.432 | 0.084 | 0.378 | 0.435 | 0.488 | 0.057 | 0.054 |
| Physiol | 808 | 0.443 | 0.088 | 0.385 | 0.443 | 0.496 | 0.058 | 0.053 |
| Gastro | 1005 | 0.464 | 0.095 | 0.402 | 0.466 | 0.528 | 0.064 | 0.062 |

(b) For 12 journals

| Journal | $n$ | Mean | S.D. | P25 | P50 | P75 | P50−P25 | P75−P50 |
|---|---|---|---|---|---|---|---|---|
| CondMat-1 | 570 | 0.437 | 0.130 | 0.352 | 0.436 | 0.531 | 0.084 | 0.095 |
| CondMat-2 | 4046 | 0.417 | 0.115 | 0.337 | 0.417 | 0.491 | 0.080 | 0.074 |
| Inorg-1 | 882 | 0.458 | 0.098 | 0.397 | 0.458 | 0.526 | 0.062 | 0.068 |
| Inorg-2 | 626 | 0.438 | 0.099 | 0.374 | 0.437 | 0.504 | 0.064 | 0.066 |
| Elec-1 | 228 | 0.460 | 0.110 | 0.380 | 0.469 | 0.535 | 0.089 | 0.066 |
| Elec-2 | 97 | 0.495 | 0.138 | 0.411 | 0.490 | 0.588 | 0.079 | 0.099 |
| Biochem-1 | 754 | 0.448 | 0.100 | 0.383 | 0.454 | 0.513 | 0.071 | 0.059 |
| Biochem-2 | 5469 | 0.430 | 0.081 | 0.377 | 0.433 | 0.484 | 0.055 | 0.051 |
| Physiol-1 | 107 | 0.430 | 0.072 | 0.389 | 0.423 | 0.475 | 0.034 | 0.052 |
| Physiol-2 | 459 | 0.448 | 0.088 | 0.390 | 0.448 | 0.500 | 0.057 | 0.052 |
| Gastro-1 | 385 | 0.471 | 0.100 | 0.403 | 0.478 | 0.544 | 0.075 | 0.066 |
| Gastro-2 | 269 | 0.468 | 0.094 | 0.405 | 0.472 | 0.535 | 0.068 | 0.063 |
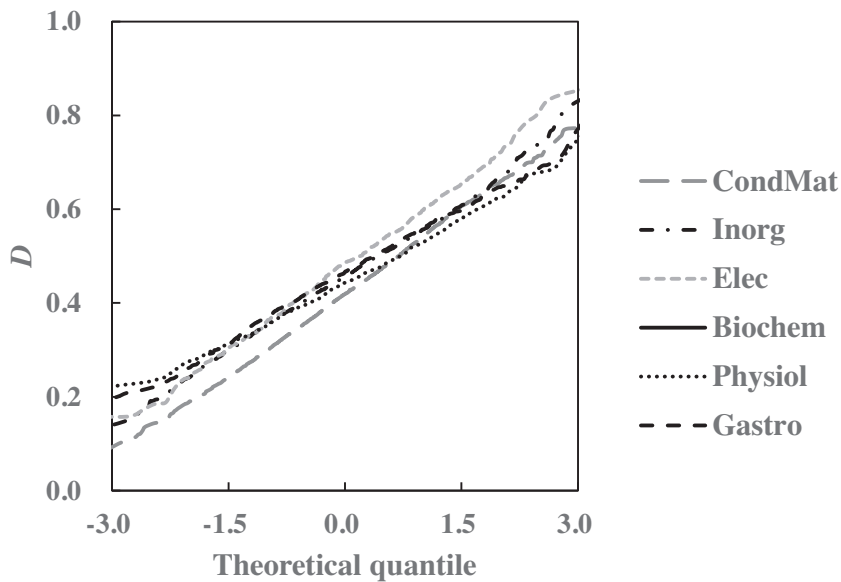


**Fig. 4.** Q-Q plots for $D$ of the articles with $C(T) \geqq 1$.

### 4.2.3. Differences of mean D for different fields

It is well known that citation distribution of articles largely differs from field to field; Table 4 and Table 5 also demonstrate this tendency. Does this hold for the distribution of $D$? Distributions of $D$ and $\log[C(T)+1]$ (both for articles $C(T) \geqq 1$) are shown using box plots in Fig. 5a and Fig. 5b, respectively, where the difference among fields can be visually confirmed. Compared to $\log[C(T)+1]$, the distribution of $D$ seems to considerably overlap among the different fields.

To confirm whether that is true or not, overlaps between distributions of two fields were calculated for all field pairs for both $D$ and $\log[C(T)+1]$ by the following methods.:

(1) First, for each field A, 25 and 75 percentile values (P25 and P75) were identified (for both $D$ and $\log[C(T)+1]$).
(2) Next, for each field B other than A, the interval of percentile values overlapping the interval [P25, P75] of field A was calculated.
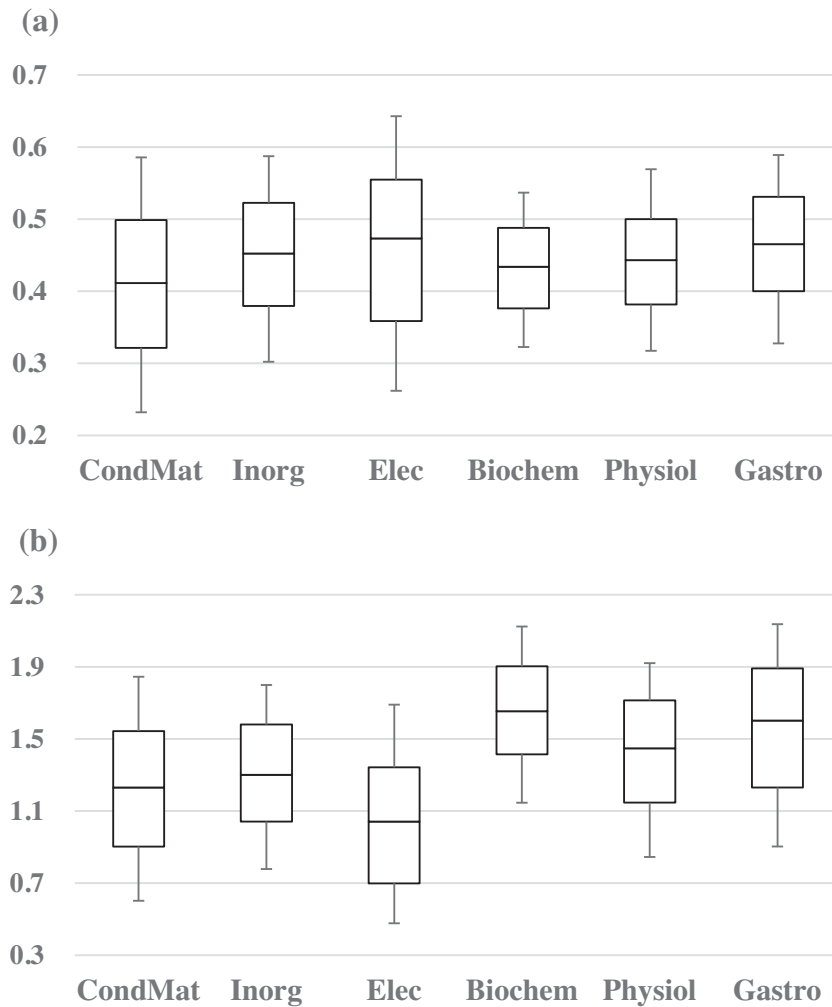
**Fig. 5.** Difference among fields of (a) distribution of $D$ and (b) distribution of $\log[C(T)+1]$.
The bar in the middle of the box is the median (the second quartile), the lower and upper boundary of the box indicate the first and third quartile, respectively, and the lower and upper bar outside the box are the 10 and 90 percentile, respectively.

For example, in case 'CondMat' is field A, its [P25, P75] of $D$ is [0.321, 0.499]. For the distribution of $D$ in 'Inorg' (field B), the values 0.321 and 0.499 correspond to 12.7 and 66.1 percentiles, respectively. Therefore, the overlap of percentile interval of field B to field A is 53.4 in this case.

These overlaps are obtained for 30 pairs (6 field A's times 5 field B's). The cumulative distributions of the overlaps for $D$ and $\log[C(T)+1]$ shown in Fig. 6 demonstrate that overlapping among the fields of $D$ is obviously larger than that of $\log[C(T)+1]$.

These findings indicate that the citation durability is not so much dependent on the subject field as the citation count. Although many previous studies have revealed that the citation count level among fields follows the general trend of biomedical > physical/chemical > engineering, $D$ does not show such a trend.

### 4.3. Relationship between D and log[C(T) + 1]

Table 8 shows Pearson's product-moment correlation coefficients $r(D\text{-}C)$ between $D$ and $\log[C(T)+1]$ calculated for the articles with $C(T) \geqq 5$. All coefficients lie in a range of 0.15–0.5, and do not include 0 in their 95% confidence intervals. Spearman's rank correlation coefficients had similar values. This suggests that a greater total number of citations means a higher citation durability.

A closer examination reveals, however, that the relationship is not linear. Table 9 shows the mean of $\log[C(T)+1]$ values for $D$ in intervals of 0.05 for each field and indicates that $\log[C(T)+1]$ assumes a maximum at a certain value of $D$. This is illustrated as scatter diagrams for the fields 'CondMaT' and 'Biochem' in Fig. 7. The mean values (dots in the figures) show nonlinear relationships of $\log[C(T)+1]$ with $D$. In other words, citations of a frequently cited article tend to spread over the entire period instead of being concentrated to very early or very late phases.
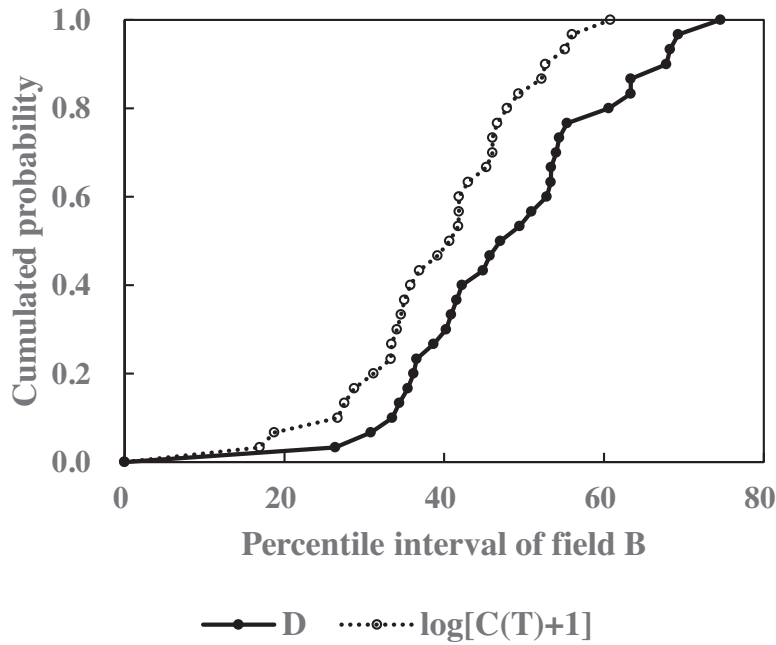
**Fig. 6.** Percentile intervals of field B overlapping the interval [P25, P75] of field A.

**Table 8**
Pearson's correlation coefficients between $D$ and $\log[C(T)+1]$, $r(D\text{-}C)$.

(a) For 6 subject fields

| Field | $n$ | $r(D\text{-}C)$ | 95% CI |
|---|---|---|---|
| CondMat | 5097 | 0.325 | [0.300, 0.350] |
| Inorg | 2061 | 0.167 | [0.124, 0.208] |
| Elec | 517 | 0.259 | [0.177, 0.338] |
| Biochem | 7327 | 0.288 | [0.267, 0.309] |
| Physiol | 808 | 0.291 | [0.227, 0.353] |
| Gastro | 1005 | 0.198 | [0.138, 0.256] |

(b) For 12 journals

| Journal | $n$ | $r(D\text{-}C)$ | 95% CI |
|---|---|---|---|
| CondMat-1 | 570 | 0.330 | [0.254, 0.401] |
| CondMat-2 | 4046 | 0.366 | [0.339, 0.392] |
| Inorg-1 | 882 | 0.271 | [0.209, 0.331] |
| Inorg-2 | 626 | 0.159 | [0.082, 0.235] |
| Elec-1 | 228 | 0.260 | [0.135, 0.377] |
| Elec-2 | 97 | 0.321 | [0.130, 0.489] |
| Biochem-1 | 754 | 0.323 | [0.258, 0.386] |
| Biochem-2 | 5469 | 0.324 | [0.300, 0.348] |
| Physiol-1 | 107 | 0.471 | [0.309, 0.607] |
| Physiol-2 | 459 | 0.346 | [0.263, 0.424] |
| Gastro-1 | 385 | 0.293 | [0.199, 0.382] |
| Gastro-2 | 269 | 0.210 | [0.092, 0.321] |

It is seen in Fig. 7 that the range of $D$ is widely spread for lower citation counts (in the lower part of the graph) but becomes narrower for higher counts (in the upper part). This is consistent with what was stated in Subsection 4.2.1, i.e. the distribution of $D$ is narrower if only highly cited papers are considered.

It should be noted that the mean $D$ values for $\log[C(T)+1]$ intervals form monotonically increasing curves, not showing a peak at a certain value of $\log[C(T)+1]$.

### 4.4. Citation durability in different journals

In the preceding subsection it was shown that an article with higher citations tends to have higher citation durability. Then, do articles in a high impact journal also have high citation durability in average? Table 10 shows the means of $D$ (of the

**Table 9**
The relationship between $D$ and $\log[C(T)+1]$.

| Range of $D$ | Mean of $\log[C(T)+1]$[a] | | | | | |
|---|---|---|---|---|---|---|
| | CondMat | Inorg | Elec | Biochem | Physiol | Gastro |
| $0.05 \leq D < 0.1$ | 0.88 | | | 0.78 | | |
| $0.1 \leq D < 0.15$ | 0.96 | 0.92 | | 0.95 | | |
| $0.15 \leq D < 0.2$ | 1.06 | 0.99 | 0.97 | 1.11 | | 1.02 |
| $0.2 \leq D < 0.25$ | 1.08 | 1.07 | 0.97 | 1.21 | 1.08 | 1.17 |
| $0.25 \leq D < 0.3$ | 1.18 | 1.14 | 1.00 | 1.34 | 1.24 | 1.26 |
| $0.3 \leq D < 0.35$ | 1.25 | 1.28 | 1.07 | 1.49 | 1.31 | 1.44 |
| $0.35 \leq D < 0.4$ | 1.34 | 1.35 | 1.15 | 1.63 | 1.42 | 1.57 |
| $0.4 \leq D < 0.45$ | 1.42 | 1.42 | 1.19 | 1.69 | 1.48 | 1.62 |
| $0.45 \leq D < 0.5$ | 1.47 | 1.45 | 1.39 | 1.75 | 1.57 | 1.69 |
| $0.5 \leq D < 0.55$ | 1.49 | 1.42 | 1.29 | 1.75 | 1.60 | 1.66 |
| $0.55 \leq D < 0.6$ | 1.53 | 1.46 | 1.37 | 1.77 | 1.63 | 1.71 |
| $0.6 \leq D < 0.65$ | 1.48 | 1.39 | 1.36 | 1.71 | 1.59 | 1.70 |
| $0.65 \leq D < 0.7$ | 1.50 | 1.31 | 1.41 | 1.56 | 1.10 | 1.50 |
| $0.7 \leq D < 0.75$ | 1.63 | 1.26 | 1.43 | 1.55 | | 1.59 |
| $0.75 \leq D < 0.8$ | 1.13 | 0.91 | 0.93 | 0.85 | | |
| $0.8 \leq D < 0.85$ | 0.86 | 0.88 | 1.00 | | | |
| $0.85 \leq D < 0.9$ | 1.18 | | | | | |

[a] The values are shown only in the case the number of articles included is equal to or larger than 2.
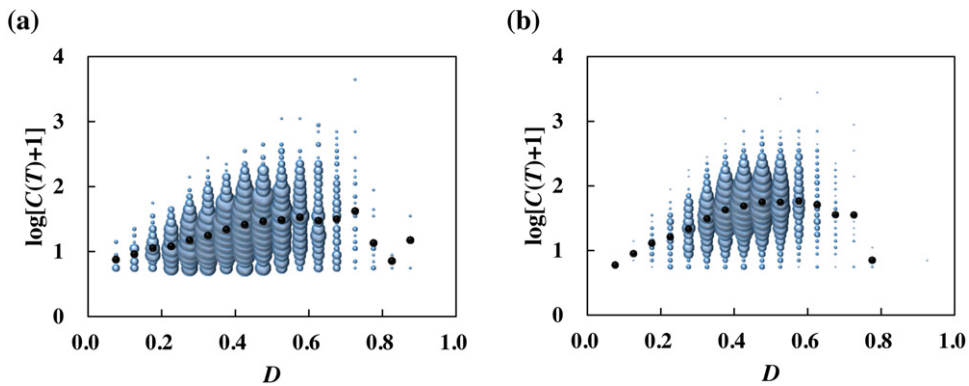


**Fig. 7.** Non-linear relations of $\log[C(T)+1]$ with $D$ for (a) the 'CondMat' field and (b) the 'Biochem' field.
Bubble sizes show the number of articles falling under each division. Dots (●) shows the mean of $\log[C(T)+1]$ in each $D$ region.

**Table 10**
Pearson's correlation coefficients ($r$) between the mean $D$ of journal and the journal impact factor ($IF$) in each field.

| Field | Journal | $D$ mean | $IF$ | $r$ |
|---|---|---|---|---|
| CondMat | Eur Phys J B | 0.411 | 1.78 | −0.872 |
| | J Phys-Condens Matter | 0.422 | 1.69 | |
| | Physica B | 0.444 | 0.64 | |
| | Phys Rev B | 0.408 | 3.20 | |
| Inorg | Inorg Chem | 0.453 | 2.95 | −0.873 |
| | Inorg Chim Acta | 0.460 | 1.48 | |
| | J Chem Soc-Dalton Trans | 0.434 | 2.92 | |
| | Transit Met Chem | 0.476 | 0.86 | |
| Elec | IEE Proc-Circuit Device Syst | 0.437 | 0.24 | 0.185 |
| | IEEE Trans Circuits Syst I | 0.488 | 0.80 | |
| | IEEE Trans Microw Theory Tech | 0.449 | 1.39 | |
| | Signal Process | 0.459 | 0.51 | |
| Biochem | Eur J Biochem | 0.443 | 2.92 | −0.298 |
| | J Biochem (Tokyo) | 0.428 | 1.93 | |
| | J Biol Chem | 0.430 | 6.98 | |
| | J Mol Biol | 0.430 | 5.59 | |
| Physiol | J Gen Physiol | 0.432 | 5.34 | −0.487 |
| | Jpn J Physiol | 0.459 | 1.10 | |
| | J Physiol-London | 0.444 | 4.56 | |
| | Pflugers Arch | 0.432 | 1.66 | |
| Gastro | Am J Gastroenterol | 0.468 | 3.75 | −0.998 |
| | Gastroenterology | 0.446 | 13.23 | |
| | Gut | 0.463 | 6.25 | |
| | J Gastroenterol | 0.473 | 1.35 | |

**Table 11**
Pearson's correlation coefficients between $D$ and the explanatory variables.

| Variable | CondMat | Inorg | Elec | Biochem | Physiol | Gastro |
|---|---|---|---|---|---|---|
| Authors | −0.019 | −0.092 | −0.001 | 0.027 | 0.020 | **0.106** |
| Insts | −0.027 | ***−0.143*** | 0.042 | −0.009 | 0.025 | 0.073 |
| Countries | −0.059 | −0.085 | 0.036 | 0.023 | −0.025 | **0.171** |
| Refs | −0.009 | −0.097 | 0.016 | 0.019 | −0.068 | ***−0.101*** |
| Price | ***−0.102*** | ***−0.179*** | −0.037 | ***−0.134*** | −0.082 | ***−0.325*** |
| Length | 0.062 | −0.045 | 0.043 | 0.030 | −0.087 | −0.028 |
| Figures | 0.006 | −0.043 | ***−0.119*** | ***−0.115*** | ***−0.109*** | −0.025 |
| Tables | **0.105** | **0.152** | **0.128** | **0.126** | 0.018 | **0.284** |
| Eqs | 0.057 | 0.004 | 0.028 | **0.101** | ***−0.164*** | – |
| Age | −0.036 | ***−0.123*** | 0.007 | 0.034 | −0.074 | 0.001 |
| RatePubl | **0.163** | −0.024 | 0.082 | −0.026 | −0.005 | ***−0.101*** |
| MedCite | −0.056 | −0.067 | ***−0.173*** | 0.033 | 0.062 | −0.064 |
| C(T) | **0.222** | **0.104** | **0.300** | **0.198** | **0.185** | **0.287** |

Bold letters: $r > 0.1$ Bold italic letters: $r < −0.1$.

**Table 12**
Pearson's correlation coefficients between each pair of the explanatory variables.

| | Authors | Insts | Countries | Price | Refs | Length | Figures | Tables | Eqs | Age | RatePubl | MedCite | C(T) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Authors* | | 6 | 6 | 3 | 1 | 2 | 2 | 2 | | 1 | | | 2 |
| *Insts* | | | 6 | 1 | 1 | 2 | 1 | 1 | | 2 | | 1 | 3 |
| *Countries* | | | | 1 | 1 | | | | | | | 1 | 1 |
| *Price* | | | | | | | 1 | | | | 1 | 2 | 6 |
| *Refs* | 1 | | | | | 6 | 5 | 2 | 3 | | 1 | 1 | 5 |
| *Length* | 1 | 1 | | | | | 6 | 4 | 5 | | 1 | 1 | 4 |
| *Figures* | | 1 | | | | | | 1 | 2 | | | | 2 |
| *Tables* | | | | 2 | | 1 | | | 1 | | | | 1 |
| *Eqs* | 2 | | | 1 | | | | | | | | | |
| *Age* | | | | 3 | 1 | | | | | | 6 | 6 | |
| *RatePubl* | | | | 1 | | | | | | | | 3 | 2 |
| *MedCite* | | | | | | | | | 1 | | | | |
| *C(T)* | | | | 1 | | | | | | 1 | | | |

Figures in upper left and lower right triangles mean the numbers of fields for which $r > 0.15$ and $r < −0.15$, respectively.

articles with $C(T) \geqq 1$) for four journals each of different fields along with the impact factor *IF* of each journal. The *IF* shown is the mean of the values for 2001 and 2002 since the articles in question were published in 2000.

As seen from the table, Pearson's correlation coefficients between $D$ mean and *IF* are negative for all the fields except 'Elec'; the correlation is considerably strong. Although its significance is not determined due to the small number of journals set at four, it does not seem to be mere coincidence. This issue is further discussed later in Subsection 4.5.2 and Subsection 5.2.

### 4.5. Linear multiple regression analysis on relationships of the citation durability with other characteristics of articles

As stated in Subsection 3.5, multiple regression analysis was performed on relatively small samples (see Table 3 for the sizes for individual subject fields) extracted from articles with $C(T) \geqq 5$.

#### 4.5.1. Preliminary analysis: correlation between variables

Prior to executing multiple regression analysis, Pearson's correlation coefficients $r$ between the response variable $D$ and individual explanatory variables were calculated as shown in Table 11. $C(T)$ and *Tables* are positively correlated with $D$ in most fields, with $r > 0.1$ in all (six) and five fields, respectively. On the other hand, *Price* and *Figures* tend to be negatively correlated with $D$, showing $r < −0.1$ in four and three fields, respectively. Hence, these four variables are expected to show the significant relation with $D$ in multiple regression analysis although the correlations are not so strong (absolute values of $r$ not exceeding 0.3). Fig. 8 shows the $r$ values and their 95% confidence intervals for these four variables.

Correlations between the explanatory variables are shown in Table 12. The numerals in this table are the number of the fields in which the absolute values of the Pearson correlation coefficient is larger than 0.15; positive and negative
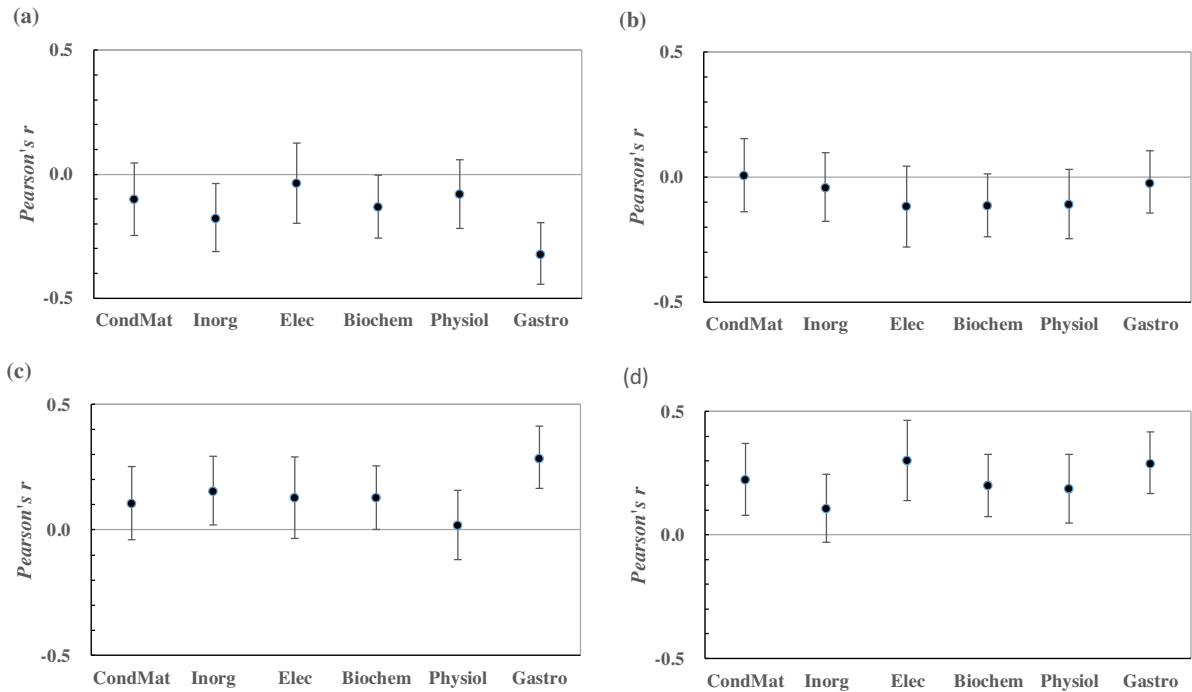
**Fig. 8.** Pearson's correlation coefficients of $D$ with (a) *Price*, (b) *Figures*, (c) *Tables*, and (d) $\log[C(T)+1]$ and their 95% confidence intervals. The bars show 95% confidence intervals.

**Table 13**
The results of the multiple regression analysis: the coefficient of determination ($R^2$), adjusted R squared ($R_c^2$), and variance ratio ($F$).

|           | CondMat | Inorg | Elec  | Biochem | Physiol | Gastro |
|-----------|---------|-------|-------|---------|---------|--------|
| $R^2$     | 0.136   | 0.131 | 0.177 | 0.153   | 0.130   | 0.350  |
| $R_c^2$   | 0.116   | 0.108 | 0.154 | 0.138   | 0.112   | 0.330  |
| $F$       | 6.76    | 5.61  | 7.54  | 10.20   | 7.23    | 17.23  |

correlations in the upper right and lower left triangles, respectively. Groups of variables showing relatively high correlation were *Authors-Insts-Countries*, *Refs-Length-Figures* and *Age-RatePubl-MedCite*. $C(T)$ shows positive correlations with several variables. Variable pairs *Authors-Insts*, *Insts-Countries*, *Refs-Length* and *Length-Figures* show absolute values of $r$ over 0.5 in three fields or more, which, however, seldom exceed 0.7. Consequently, it was concluded that the multicollinearity problem can be ignored in the present case; the multiple regression analysis involved all the variables, and significant variables were chosen by the stepwise technique.

### 4.5.2. Multiple regression analysis

Multiple regression analysis with $D$ as the response variable and the characteristics shown in Table 2 as the explanatory variables resulted in significant correlations (with probability of significance $p$ less than $10^{-4}$) in all the six fields. Coefficients of determination $R^2$, adjusted coefficients of determination $R_c^2$ and variance ratios $F$ are summarized in Table 13. $R_c^2$ values are not particularly high and lie in a range of 0.1–0.3.

Table 14 shows the standardized partial correlation coefficients β and their 95% confidence intervals for the explanatory variables. The criteria for selection of significant variables in the stepwise approach were 10% as the upper limit for selection and 20% as the lower limit for elimination. The table therefore indicates β's of variables that are significant at 10% level, hence there are a few β's that include 0 in the 95% confidence interval. The relatively small sample size (about 200) was the reason for setting the moderate level of significance. In the field 'Gastro', the variable *Eqs* (number of equations) was not included in the analysis because none of the articles in this field included equations.

The following five variables were selected at least in three subject fields out of six:

(1) *Price*

Negatively correlated with $D$ in five subject fields other than 'Elec', meaning that articles that cite more recent references (within five years after publication) tend to be cited earlier.

(2) *Figures* and *Tables*

Positive and negative correlation, respectively, with $D$ in three fields each. This means that articles containing more figures tend to be cited earlier, and those with more tables later. These relationships are, however, not so strong.

**Table 14**
The results of the multiple regression analysis: the standardized partial regression coefficients (β). [a,b]

| Variable | CondMat | Inorg | Elec | Biochem | Physiol | Gastro |
|---|---|---|---|---|---|---|
| Authors | | | | | | |
| Insts | | −0.164 ± 0.139 | | | | |
| Countries | | | | | | .136 ± 0.122 |
| Refs | | | | | | −0.163 ± 0.132 |
| Price | −0.184 ± 0.149 | −0.223 ± 0.140 | | −0.302 ± 0.143 | −0.232 ± 0.145 | −0.380 ± 0.123 |
| Length | | −0.185 ± 0.171 | | | | |
| Figures | | | *−0.128 ± 0.152* | −0.136 ± 0.125 | −0.207 ± 0.157 | |
| Tables | | 0.226 ± 0.154 | 0.185 ± 0.153 | | | *.111 ± 0.125* |
| Eqs | | | | | *−0.136 ± 0.143* | |
| Age | | | | | | |
| RatePubl | 0.160 ± 0.142 | | | | | |
| MedCite | | | −0.212 ± 0.152 | | | |
| C(T) | 0.308 ± 0.152 | 0.241 ± 0.153 | 0.340 ± 0.154 | 0.423 ± 0.148 | 0.354 ± 0.156 | .464 ± 0.150 |
| IF | −0.183 ± 0.147 | | | −0.144 ± 0.141 | | −0.191 ± 0.154 |

[a] The estimated value and 95% confidence interval of coefficients are shown if selected in a stepwise process.
[b] Italic letters: contain 0 in the 95% confidence interval.

**Table 15**
Pearson's correlation coefficients $r(D\text{-}H)$ and Kendall's rank correlation coefficients $\tau_b(D-H)$ between $D$ and $H$ for 6 subject fields.

| Field | $n$ | $r(D\text{-}H)$ | $\tau_b(D\text{-}H)$ |
|---|---|---|---|
| CondMat | 5097 | 0.911 | 0.784 |
| Inorg | 2061 | 0.909 | 0.781 |
| Elec | 517 | 0.918 | 0.790 |
| Biochem | 7327 | 0.918 | 0.800 |
| Physiol | 808 | 0.905 | 0.775 |
| Gastro | 1005 | 0.927 | 0.818 |

(3) $C(T)$ and $IF$

$D$ has relatively strong positive correlation with $C(T)$ in all the fields and somewhat weak negative correlation with $IF$ in three fields. This result is consistent with the result of correlation analysis for all $C(T) \geqq 5$ articles (see Subsection 4.3 and Subsection 4.4).

The multiple regression analysis results thus underlines the results of the preliminary correlation analysis: it revealed more definitive relation of $D$ with $C(T)$, $Price$, $Figures$, and $Tables$ than the correlation analysis in which those relations were somewhat obscure.

Correlation with the variables other than the five variables mentioned above was not significant.

The sample in each field consists of nearly the same number of articles extracted from four journals each. To check for the possibility of a bias by the journals, multiple regression analysis was performed on each of the 24 journals. $D$ was found to be significantly correlated with $C(T)$, $Price$, $Figures$ and $Tables$ for 15, 7, 2 and 5 journals, respectively, which is consistent with the results of analysis by field, although the small sample size limits the explanatory power.

## 5. Discussion

### 5.1. Relationship between D and median citation age H

The median citation age $H$ is sometimes used as a measure of citation durability because it can be calculated easily. The $H$ of an article is the first year $t$ where the relative cumulative citation count $x(t)$ defined by Eq. (2) becomes 0.5 or more. Here, some properties of $H$ are compared with those of $D$.

(1) Distribution of $H$

Fig. 9 shows Q-Q plots that compare the distribution of $H$ for articles with $C(T) \geqq 1$ to normal ones in the six fields. Comparison of Fig. 9 with Fig. 4 reveals that the plots for $H$ are not as linear as those for $D$, indicating that the distributions of $H$ are not closer to normal than those of $D$.

(2) Relationship of $H$ with $D$

Table 15 shows the Pearson correlation coefficients between $D$ and $H$, $r(D\text{-}H)$, for $C(T) \geqq 5$ articles. The correlation is very high with $r(D\text{-}H) > 0.9$ for all fields. This could suggest that the citation durability may be measured by $H$ rather than $D$ because the former is easier to calculate. However, even two measures that have a high correlation often yield the considerably different rankings. The Kendall rank correlation coefficients $\tau_b$, also shown in Table 15, indicate the extent of discrepancy between the two rankings. In $n(n-1)/2$ pairs from $n$ data in the sample, let the ratio of identically ranked pairs be $y$; then the ratio of reversely ranked pairs is $1-y$. By definition, $\tau_b = y-(1-y)$, then $y = (1 + \tau_b)/2$. (Though there is the problem of tie data, let us neglect it to make the explanation simpler.) The range of 0.78–0.82 of $\tau_b$ in Table 15 means $y = 0.89$–0.91. In other words, about 10% of the data pairs receive reverse ranking by $H$ and $D$.
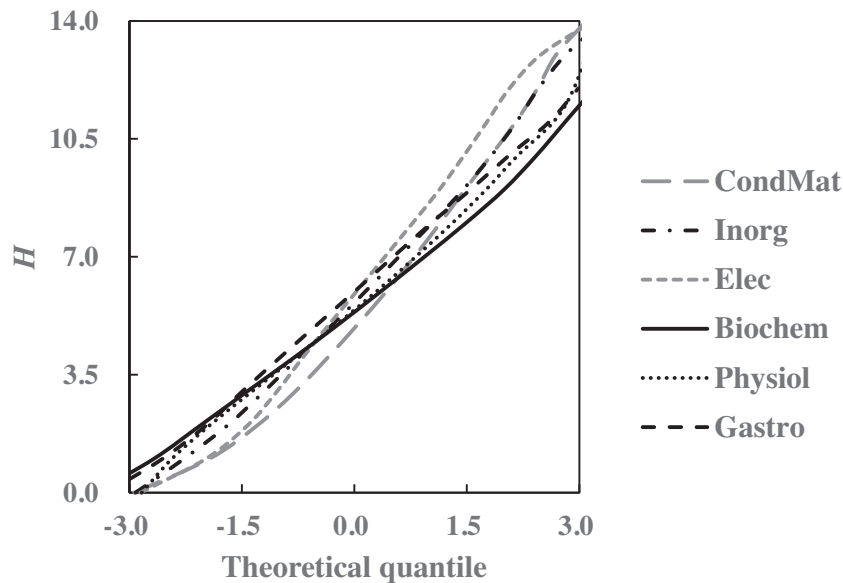
**Fig. 9.** Q-Q plots for *H* of the articles with $C(T) \geqq 1$.

While *D* is based on the overall information of the citation age distribution, *H* is related only to the median of the distribution. This is the reason why they yield different rankings of articles in spite of high correlation coefficients. Consequently, *H* is not preferable as a measure of citation durability. Another advantage of *D* is its distribution with higher normality than that of *H*.

### 5.2. Relationship of D with the total number of citations and with the impact factor

The following results were demonstrated in the present work:

(1) *D* is positively correlated with *C*(*T*) (see Table 8). Multiple regression analysis also showed that the partial regression coefficient of *C*(*T*) was significantly positive in all the fields (see Table 14).
(2) However, *C*(*T*) does not linearly increase with *D*, but the mean of *C*(*T*) has a maximum value at a certain value of *D* (see Table 9 and Fig. 7).
(3) *D* tends to be rather low for journals with high impact factors (*IF*) (see Table 10). In multiple regression analysis, the partial regression coefficient of *IF* was significantly negative in three fields out of six (see Table 14).

As for (1) above, several studies have reported that more cited papers in the long run tend to receive their citations more lately than fewer cited ones (Aversa 1985; Costas et al., 2010; Levitt & Thelwall 2008; Levitt & Thelwall 2009; Line 1984; Wang 2013). This work is the first to verify this relation by multiple regression analysis using many controlling variables over six different fields

Our results concerning (2) above can be compared with those of Costas et al. (2010) who examined the field-normalized citation counts (*CPP/FCSm*) among three classes of different citation durability: early cited [II], delayed cited [III], and normal [IV]. For citation windows of five, ten and twenty years after publication, the orders of mean *CPP/FCSm* were IV > II > III, IV > III > II, and III > IV > II, respectively. In the present work using the citation window of 15 years, though there is an increasing tendency of *C*(*T*) with *D*, the mean *C*(*T*) has a maximum values at a certain value of *D*. This is consistent with the trend seen for the 10-year window by Costas et al.

The relation shown in (3) above is reported for the first time in the present work. This means that more highly cited papers tend to have higher citation durability in individual journals while journals with high *IF* tend to include more papers with lower citation durability in higher proportions. This seemingly contradictory result may lead us to an idea that it is due to the fact that *IF* considers only short-term citations (within two years after publication). However, the considerably strong negative correlations are also found between the journal mean of *D* and the journal mean of *C*(*T*), which is a long-term (for 15 years after publication) citation count, in most fields, as follows:

CondMat $r = -0.812$
Inorg $r = -0.756$
Elec $r = 0.670$
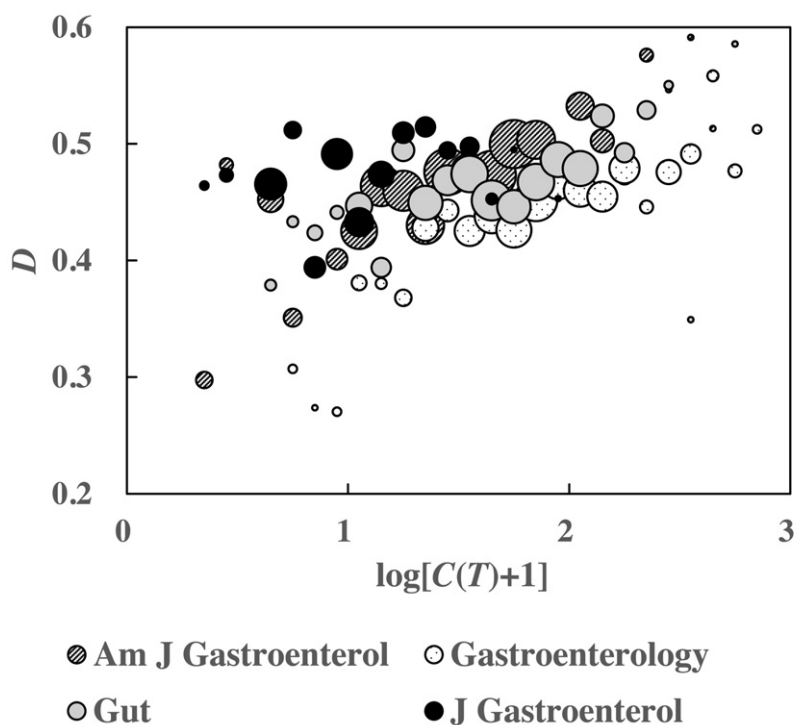Biochem $r = -0.248$
Physiol $r = -0.499$

**Fig. 10.** Comparison of $C(T)$-$D$ distributions among journals in the 'Gastro' field.
Each bubble represents the mean $D$ value of articles falling under the individual $\log[C(T)+1]$ division. Bubble sizes show the number of articles included in the division.

Gastro $r = -0.975$

These results support the supposition that a journal of higher citation impact, whether it is short-ranged or long-ranged, contains articles with lower citation durability at a higher ratio, although not claimed definitely because there is one exception ('Elec' field) among the six fields. Fig. 10 illustrates this situation in the case of 'Gastro' field; the centers of the distributions within journals shift from lower right to upper left, being arranged in order of Gastroenterology (the lowest $D$ and the highest $C(T)$), Gut, Am J Gastroenterol, and J Gastroenterol (the highest $D$ and the lowest $C(T)$). However, the whole tendency (and also within the individual journals) shows an increase of $D$ when $C(T)$ increases.

### 5.3. Relationships of D with other characteristics of articles

As discussed in Subsection 2.5, studies of citation durability in relation to article characteristics are limited to those by van Dalen and Henkens (2005), Costas et al. (2010), and Wang et al. (2015). Their results are compared with the present work in the following.

van Dalen and Henkens (2005) classified the 1371 articles published in 17 demographic journals in 1990–92 according to the citation pattern in 10 years after publication into four classes: [I] few cited, [II] early cited, [III] delayed cited, and [IV] normal. They discussed effects of the characteristics of individual articles on their attribution to the classes using logit polynomial regression analysis with the class I as the reference. Among the explanatory variables, publication in top-ranked journals, author's reputation (higher citation count in the past), and number of pages were found to increase the probability for the article to belong to the class II, III or IV compared to the class I; number of authors and author's national affiliation were not significant. Although their results cannot be compared with the present work directly due to the lack of mutual comparison among the classes II, III and IV, the data suggests that the three significant variables are correlated most strongly with the class IV. As for the classes II and III, publication in top-ranked journals favors allocation to the class II, but the other two variables show no difference.

Similar classification into the four classes (but with a different criterion) was performed by Costas et al. (2010) for articles published in 1983–2003 and indexed in WoS according to citation data up to 2008. Mutual comparison of the classes II, III and IV revealed that the articles in the class III have significantly smaller number of authors, number of affiliated institutions, number of national affiliations and number of cited references, and greater number of pages than those in the class II.

Using the citation data of 13 years received by articles published in 2001, Wang et al. (2015) carried out multiple regression analysis with Citation Delay ($D$ in this paper) as the dependent variables and several measures of interdisciplinarity of articles

**Table 16**

Comparison between the previous and present studies on the significance of correlation of various article features with the citation durability.

| Feature | van Dalen and Henkens (2005) | Costas et al. (2010) | Wang et al. (2015) | Present work |
|---|---|---|---|---|
| #Authors | Not | Negative | Negative | Not |
| #Institutions | | Negative | | Not |
| #Countries | | Negative | Negative | Not |
| #References | | Negative | Negative | Not |
| Article length | Not | Positive | Positive | Not |
| Author performance | Not | | | Not |
| Journal prestage | Negative | | | Negative |
| Total citations | | Positive | | Positive |
| Price index | | | | Negative |
| #Figures | | | | Negative |
| #Tables | | | | Positive |

(measured by the distribution of disciplines to which their cited references belong) and also some controlling variables as the independent variables. Their main interest lay in the relationship between interdisciplinarity and $D$, but I address here the controlling variables that are more general article characteristics. Wang et al. showed that, among the controlling variables, number of authors, internationality of co-authors, and number of references had significant negative effect on $D$ while number of pages had significant positive effect.

The results of the three studies mentioned above are compared to those of the present work in Table 16. The results of van Dalen and Henkens are consistent with what our multiple regression analysis shows: citation durability has no definite relationship with *Authors*, *Length* and *MedCite* (corresponding to number of authors, length of the article, and author's reputation of van Dalen and Henkens, respectively) but is negatively correlated with *IF* (corresponding to publication in top-ranked journals). On the contrary, Costas et al. and Wang et al. suggested significant relationships of citation durability with the degree of collaboration, number of references, and article length, which are not recognized as significant in this work. The reasons of the discrepancy are considered as:

1. Although the methodology used by Wang et al. is to some extent similar to that used by this study (both detecting characteristics having the significant relation to citation durability by multiple regression analysis), approaches by van Dalen and Henkens and also by Costas et al. are largely different. Their studies focused more on describing main differences about the typologies of durability and used different statistical tests as those used in this study.
2. The sample sizes used are greatly different. Those used in this study and the study of van Dalen and Henkens are of the order of $10^2$ or $10^3$, while Costas et al. used about 8 million and Wang et al. used 0.3 million articles. This leads to considerably different effect sizes to be statistically significant (although it is impossible to compare the effect sizes because different analytic methods are adopted by the respective studies).

The significant relationships of citation durability with some variables (the degree of collaboration, number of references, and article length) were recognized in the studies by Costas et al. and Wang et al. while not detected in this work. It might be because these relationships were controlled by other variables used in multiple regression analysis in this work. In order to examine this possibility, I tried to carry out multiple regression analysis excluding $C(T)$, which is the variables most strongly related with $D$, from the explanatory variables. Contrary to expectation, when $C(T)$ was excluded, the cases resulted in a rise in significance or an increase in the effect size (standardized partial correlation coefficient) were much fewer than the opposite cases. The variables such as *Authors*, *Insts*, *Refs*, and *Length* were still not significant in most fields, while *Price*, selected in 5 fields as 10% significant in regression including $C(T)$, was selected in only 3 fields when excluding $C(T)$. Similarly, both *Figure* and *IF* reduced the selected fields from 3 to 1 by excluding $C(T)$, These results suggest that negative correlations between $D$ and these variables are revealed more definitely by the controlling variable $C(T)$.

### 5.4. Comparison with characteristics that affect citation counts

The authors have reported characteristics of articles that affect citation counts (Onodera & Yoshikane 2015) based on practically the same sample articles, and using the same explanatory variables except $C(T)$ and *IF* (a dummy variable for journals instead of *IF*) as those used in the present work. These results are now drawn on for comparison between the characteristics closely related to the citation count and those closely related to the citation durability.

Variables showing relatively strong correlation with $D$ include *Price*, *Figures* and *Tables*, as described in Subsection 4.5.2 and summarized in Table 14. The correlation was negative for *Price* and *Figures*, and positive for *Tables*. On the other hand, variables related to the citation count $C$ were *Price* (the strongest correlation), *Ref* (the second strongest), and *Authors* and *MedCite* (relatively weak correlation). Correlation was positive for all the variables.

*Price* shows strong correlation with both $C$ and $D$ but with opposite signs: an article with a high *Price* value (citing more recent reference materials) will receive more total citations but with a bias to early periods. While this is not contrary to the definition of *Price*, articles with high *Price* values represent somewhat special cases, since articles with high citation counts are more likely to receive delayed citations (see Subsection 4.3).
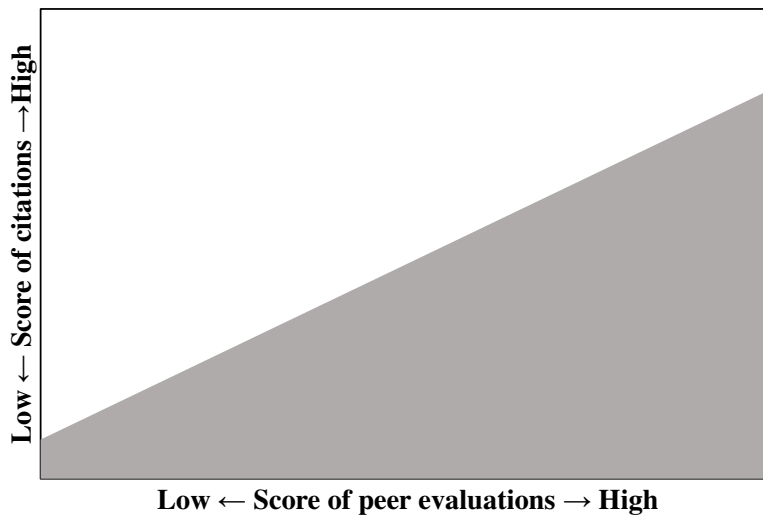
**Fig. 11.** Schematic view showing the relation between peer evaluations and citations. Most data exist in the shaded region.

*Figures* and *Tables* are also significantly correlated with *D* but their relationship with *C* was not significant. A few studies other than Onodera and Yoshikane (2015) included the number of figures and tables in an article among the explanatory variables in multiple regression analysis to explain citation counts (Haslam et al., 2008; Snizek, Oehler, & Mullins, 1991), but they also did not show a significant correlation. The present work is the first to show that the number of tables and figures in an article is correlated with the citation durability but not to the citation count. The general tendency (though not too strong) that articles containing more figures are cited earlier and those containing more tables are cited over a longer period may suggest the behavior of scholars who use more figures to attract early attention to the article while more tables for careful examination of the content in a longer period of time.

### 5.5. Relationship of article quality and citation durability

Correlation of the citation score of scientific units (papers, researchers, or research groups) with the peer review results has frequently been investigated (Abramo, D'Angelo, & Di Costa, 2011; Hayashi 2003; Mryglod, Kenna, Holovatch, & Berche, 2013; Rinia, van Leeuwen, van Vuren, & van Raan, 1998) with the intention of studying the possibility of applying bibliometric data to research evaluation practices. In most cases, these works resulted in medium correlation ($r = 0.3–0.6$), and a common observation that, while an evaluated unit with a poor score by peers receive few citations, a high score in the peer review does not necessarily mean a high citation count. The situation is schematically illustrated in Fig. 11.

The relationship between *D* and the citation count is similar to this, as shown in Fig. 7. It may suggest that citation durability is connected more closely to the quality of an article (peer review score) rather than citation count itself. Although any conclusive claim cannot be given only from this fact, it is worth testing the hypothesis using actual peer review data. It should, however, be recognized that the citation durability may not be a practical measure for research evaluation even if the hypothesis is supported, because citation durability data are available only after a lapse of considerable time.

### 5.6. On inclusion of only papers with at least 5 citations in the analysis

As described in Subsection 3.1, the present work is principally concerned with papers with $C(T)$ of 5 or more for 15 years. This is due to the following two reasons:

(a) *D* has little meaning and lacks reliability when $C(T)$ is very low.
(b) Other studies referred to in this paper also exclude few-cited papers or classified them in a different category from more cited papers.

For example, Wang et al. (2015) excluded papers with fewer than 12 citations (the median value of their data) in their multiple regression analysis "because this ratio-based measure might not be very reliable when the denominator is too small" and said "nevertheless, results are robust if we relax this restriction." Costas et al. (2010) considered in their analysis only papers with a minimum of five citations "to avoid the influence of hardly cited documents", although they classified all papers with at least one citation into three classes by their durability.

I tried a few analyses using papers with $C(T) \geq 1$ and compared the results to those obtained in the case of $C(T) \geq 5$. The correlation coefficient ($r$) between *D* and $\log[C(T)+1]$ (shown in Table 8 for the case $C(T) \geq 5$) was higher for the case $C(T) \geq 1$

in four of the six fields and in six of the twelve journals. The maximum difference between the two cases was 0.055 for the fields and 0.148 for the journals. In addition, the multiple regression analysis described in Subsection 4.5.2 was also applied to papers with $C(T) \geqq 1$. Although significant variables selected in the stepwise process somewhat decreased compared to the results in the case $C(T) \geqq 5$, no systematic change was found. The decrease of selected variables and also some deterioration of fitness (lowering of the adjusted coefficients of determination in most fields) may be due to low reliability of $D$ of few-cited papers,

From these observations, bias caused by exclusion of papers with fewer citations in the analysis is supposed to be little, if any.

### 5.7. Limitations of the present work and issues in future

As described in Subsection 4.2.3, there is some difference in the distribution of $D$ among fields although it is not so large. Therefore, it is not adequate to compare directly $D$'s of articles that belong to different fields, similarly as in the case of the citation counts. Normalization by some method is necessary for this purpose, but the issue is beyond the scope of the present work.

Many previous studies attempted to classify a paper according to the pattern of citation durability, such as a "sleeping beauty" (delayed), a "flash in the pan" (early-matured), and "normal". Although $D$ might be utilized for such classification (together with other measures), a value of $D$ does not directly correspond to one of the classes. Ke et al. (2015) who proposed "Beauty Coefficient", another indicator of citation durability, says about the distribution of the indicator "there are no clear demarcation values that allow us to separate sleeping beauties from 'normal' papers: delayed recognition occurs on a wide and continuous range." This holds in the case of $D$. Connection of quantitative measures to qualitative categories is an issue for the future.

The value of $D$ of an article varies with the observed citation period. The changing pattern would be different among articles. Therefore, the distribution shape and other properties of $D$ that are obtained from the citation history of 15 years may change when a much longer citation period is used. Or, ranking of articles according to $D$ may considerably change. They are also future issues.

Two articles whose pattern of citation history is the same have the same value of $D$ even if their total citations greatly differ. For example, if the citation count of an article has twice as many as that of another article every year, $D$ of the two articles is same. But the reverse is not always true. That is, the pattern of citation history may be quite different even if they have the same $D$ value. Recently Sun, Min, and Li (2016) introduced "obsolescence vector" as an indicator that can distinguish the different patterns to some extent. This indicator is a 2-dimensional vector consisting of a parameter $G_S$ which is proposed by Li et al. (2014) and is similar to $D$, and a parameter $A^-$ detecting drastic fluctuation of citation curves. Further development of citation durability indicators allowing to discriminate a fine difference in the structure of citation history is another issue of future investigation.

## 6. Conclusions

The main findings obtained in this study are summed up as follows according to the three objectives presented in Section 1.

**Objective 1**: Elucidating the characteristics of the distribution of "Citation Delay" ($D$).
$D$ shows the following properties:

– reflects information on the entire citation life-time of an article;
– lies between 0 and 1, getting near to 1 as citations become delayed;
– follows the normal distribution fairly closely, particularly for articles with citation counts greater than a certain level; and
– is not so much dependent on the subject field as the citation count.

These properties are favorable for the index of citation durability
**Objective 2**: Examining relationships between $D$ and the citation count of articles in different subject fields.
For all the six subject field examined, the citation data of articles over a 15-year window showed a significant positive correlation between $D$ and $\log[C(T)+1]$ (0.15–0.35 of Pearson's $r$). In addition, the standard deviation of $D$ becomes smaller for higher-cited articles. These suggest that more highly-cited articles tend to concentrate in a region of higher and also narrower citation durability. However, the relationship is not linear but the mean of $\log[C(T)+1]$ reaches a maximum at a certain value of $D$.
**Objective 3**: Examining relationships between $D$ and other characteristics of articles in different fields.
This is the first study that analyzed systematically the relationships between citation durability and other characteristics of articles. As a result of multiple regression analyses for six subject fields, several significant relationships were found.
The Price index negatively correlated with $D$. This finding together with the result of our previous study (Onodera & Yoshikane, 2015) suggests that articles that cite more recent references will receive more total citations but most of the citations are made relatively earlier.

There is a tendency, though not so strong, that articles containing more figures are cited earlier and those containing more tables are cited over a longer period.

A seemingly contradictory result is found that more highly cited papers tend to have higher citation durability in individual journals while journals with high citation impact tend to include more papers with lower citation durability in higher proportions. This tendency holds whether the citation impact of journals is measured by a short-term index (2-year impact factor) or by a long-term index (journal mean of 15-year citations).

## Author contribution

Natsuo Onodera: conceived and designed the analysis, collected the data, contributed data or analysis tools, performed the analysis and wrote the paper.

## Acknowledgments

## References

Abramo, G., Cicero, T., & D'Angelo, C. A. (2011). Assessing the varying level of impact measurement accuracy as a function of the citation window length. *Journal of Informetrics*, *5*(4), 659–667. http://dx.doi.org/10.1016/j.joi.2011.06.004

Abramo, G., D'Angelo, C. A., & Di Costa, F. (2011). National research assessment exercises: A comparison of peer review and bibliometrics rankings. *Scientometrics*, *89*(3), 929–941. http://dx.doi.org/10.1007/s11192-011-0459-x

Aksnes, D. W. (2003). Characteristics of highly cited papers. *Research Evaluation*, *12*(3), 159–170.

Aversa, E. S. (1985). Citation patterns of highly cited papers and their relationship to literature aging: A study of the working literature. *Scientometrics*, *7*(3/6), 383–389.

Avramescu, A. (1979). Actuality and obsolescence of scientific literature. *Journal of the American Society for Information Science*, *30*(5), 296–303.

Baumgartner, S. E., & Leydesdorff, L. (2014). Group-based trajectory modeling (GBTM) of citations in scholarly literature: Dynamic qualities of transient and sticky knowledge claims. *Journal of the Association for Information Science and Technology*, *65*(4), 797–811. http://dx.doi.org/10.1002/asi.23009

Burrell, Q. L. (2003). Age-specific citation rates and the Egghe–Rao function. *Information Processing and Management*, *39*(5), 761–770. http://dx.doi.org/10.1016/s0306-4573(02)00029-8

Burrell, Q. L. (2005). Are sleeping beauties to be expected? *Scientometrics*, *65*(3), 381–389.

Cano, V., & Lind, N. C. (1991). Citation life cycles of ten citation classics. *Scientometrics*, *22*(2), 297–312.

Costas, R., van Leeuwen, T. N., & van Raan, A. F. J. (2010). Is scientific literature subject to a sell-by-date? A general methodology to analyze the durability of scientific documents. *Journal of the American Society for Information Science and Technology*, *61*(2), 329–339. http://dx.doi.org/10.1002/asi.21244

Della Briotta Parolo, P., Pan, R. K., Ghosh, R., Huberman, B. A., Kaski, K., & Fortunato, S. (2015). Attention decay in science. *Journal of Informetrics*, *9*(4), 734–745. http://dx.doi.org/10.1016/j.joi.2015.07.006

Egghe, L., & Ravichandra Rao, I. K. (1992). Citation age data and the obsolescence function: Fits and explanations. *Information Processing and Management*, *28*(2), 201–217.

Garfield, E. (1980). Premature discovery or delayed recognition. Why? *Essays of an Information Scientist*, *4*, 488–493.

Garfield, E. (1984a). The 100 most-cited papers ever and how we select citation classics. *Current Contents*, *23*, 3–9.

Garfield, E. (1984b). 100 classics from the lancet. *Current Contents*, 3–13.

Glänzel, W., & Schoepflin, U. (1995). A bibliometric study on ageing and reception process of scientific literature. *Journal of Information Science*, *21*(1), 37–53.

Glänzel, W., Schlemmer, B., & Thijs, B. (2003). Better late than never? On the chance to become highly cited only beyond the standard bibliometric time horizon. *Scientometrics*, *58*(3), 571–586.

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

Hayashi, T. (2003). Is bibliometrics usefulto support the peer-review? A case study of NIAD's research evaluation in science. *Daigaku Hyoka (Research in University Evaluation)*, 167–187 (in Japanese).

Ke, Q., Ferrara, E., Radicchi, F., & Flammini, A. (2015). Defining and identifying Sleeping Beauties in science. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(24), 7426–7431. http://dx.doi.org/10.1073/pnas.1424329112

Lachance, C., & Larivière, V. (2014). On the citation lifecycle of papers with delayed recognition. *Journal of Informetrics*, *8*(4), 863–872. http://dx.doi.org/10.1016/j.joi.2014.08.002

Levitt, J. M., & Thelwall, M. (2008). Patterns of annual citations of highly cited articles and the prediction of their citation raking: A comparison across subject categories. *Scientometrics*, *77*(1), 41–60. http://dx.doi.org/10.1007/s11192-007-1946-y

Levitt, J., & Thelwall, M. (2009). Atypical citation patterns in the twenty most highly cited documents in library and information science. *Scientometrics*, *78*(1), 45–67. http://dx.doi.org/10.1007/s11192-007-1927-1

Li, J., & Shi, D. (2016). Sleeping beauties in genius work: When were they awakened? *Journal of the Association for Information Science and Technology*, *67*(2), 432–440. http://dx.doi.org/10.1002/asi.23380

Li, J., & Ye, F. Y. (2012). The phenomenon of all-elements-sleeping-beauties in scientific literature. *Scientometrics*, *92*(3), 795–799. http://dx.doi.org/10.1007/s11192-012-0643-7

Li, J., Shi, D., Zhao, S. X., & Ye, F. Y. (2014). A study of the heartbeat spectra for sleeping beauties. *Journal of Informetrics*, *8*(3), 493–502. http://dx.doi.org/10.1016/j.joi.2014.04.002

Li, J. (2014). Citation curves of all-elements-sleeping-beauties: Flash in the pan first and then delayed recognition. *Scientometrics*, *100*(2), 595–601. http://dx.doi.org/10.1007/s11192-013-1217-z

Line, M. B. (1984). Citation decay of scientific papers: Variation according to citations received. *Journal of Information Science*, *9*(2), 90–91.

Mryglod, O., Kenna, R., Holovatch, Y., & Berche, B. (2013). Absolute and specific measures of research group excellence. *Scientometrics*, *95*(1), 115–127. http://dx.doi.org/10.1007/s11192-012-0874-7

Onodera, N., & Yoshikane, F. (2015). Factors affecting citation rates of research articles. *Journal of the Association for Information Science and Technology*, *66*(4), 739–764. http://dx.doi.org/10.1002/asi.23209

Onodera, N., Iwasawa, M., Midorikawa, N., Yoshikane, F., Amano, K., Ootani, Y., et al. (2011). A method for eliminating articles by homonymous authors from the large number of articles retrieved by author search. *Journal of the American Society for Information Science and Technology*, *62*(4), 677–690. http://dx.doi.org/10.1002/asi.21491

Rinia, E. J., van Leeuwen, T. N., van Vuren, H. G., & van Raan, A. F. J. (1998). Comparative analysis of a set of bibliometric indicators and central peer review criteria. Evaluation of condensed matter physics in the Netherlands. *Research Policy*, *27*, 95–107.

Snizek, W. E., Oehler, K., & Mullins, N. C. (1991). Textual and non-textual characteristics of scientific papers. Neglected science indicators. *Scientometrics*, *20*(1), 25–35.

Sun, J., Min, C., & Li, J. (2016). A vector for measuring obsolescence of scientific articles. *Scientometrics*, *107*(2), 745–757. http://dx.doi.org/10.1007/s11192-016-1884-7

van Dalen, H. P., & Henkens, K. (2005). Signals in science—On the importance of signaling in gaining attention in science. *Scientometrics*, *64*(2), 209–233.

van Raan, A. F. J. (2004). Sleeping beauties in science. *Scientometrics*, *59*(3), 467–472.

Wang, J. (2013). Citation time window choice for research impact evaluation. *Scientometrics*, *94*(3), 851–872. http://dx.doi.org/10.1007/s11192-012-0775-9

Wang, J., Thijs, B., & Glänzel, W. (2015). Interdisciplinarity and impact: Distinct effects of variety, balance, and disparity. *PLoS One*, *10*(5), e0127298. http://dx.doi.org/10.1371/journal.pone.0127298