This is a preprint of an article accepted for publication in Journal of the American Society for Information Science and Technology. Huang, H., Andrews J., & Tang, J.(in press, 2011). Citation Characterization and Impact Normalization in Bioinformatics Journals. *Journal of the American Society for Information Science and Technology*,

# Citation Characterization and Impact Normalization in Bioinformatics Journals

Hong Huang and James Andrews

*School of Information, University of South Florida, Tampa, FL, 33620-7800.*

*Telephone: (813) 974-6361, (813) 974-6840; Fax: (813) 974-6840; E-mail: {honghuang, jimandrews}@usf.edu*

Jiang Tang

*Shanghai Information Center for Life Sciences, Chinese Academy of Sciences, Shanghai, China, 200031.*

*Telephone: (8621) 54922803; Fax: (8621)54922800; E-mail: jt@sibs.ac.cn*

## Abstract

Bioinformatics journals publish research findings of intellectual synergies among subfields such as biology, mathematics, and computer science. The objective of this study is to characterize the citation patterns in bioinformatics journals and their correspondent knowledge subfields. Our study analyzed bibliometric data (impact factor, cited-half-life, and references-per-article) of bioinformatics journals and their related subfields collected from the Journal Citation Report (JCR). The findings showed bioinformatics journals' citations are field-dependent, with scattered patterns in article life span and citing propensity. Bioinformatics journals originally derived from biology-related subfields have shorter article life spans, more citing on average, and higher impact factors. Those journals derived from mathematics and statistics demonstrate converse citation patterns. The journal impact factors were normalized, taking into account of the impacts of article life spans and citing propensity. A comparison of these normalized factors to JCR journal impact factors showed rearrangements in the ranking orders of a number of individual journals, but a high overall correlation with JCR impact factors.

## Introduction

Bioinformatics is a rapidly growing, interdisciplinary field of science that applies methods from information technology, computer science, mathematics, and statistics to solve problems in biological science (Molatudi, Molotja, & Pouris, 2009). There are a growing number of massive and heterogeneous biological data sets, including genomics, transcriptomics, and proteomics data, which were produced by new, affordable technologies like genome sequencing (Brown, 2003; Drmanac, 2010). This flood of information makes interdisciplinary teamwork using meaningful data even more challenging for bioinformatics researchers. As an independent scientific discipline, bioinformatics deals with diverse forms of data derived from many fields with different focuses. For instance, bioinformatics articles in molecular biology and medicine primarily address solving complex biological questions; bioinformatics research papers derived from computer science focus on algorithms and database developments that support biological studies; while scientists in mathematics and statistics are interested in biology theoretical modeling and quantitative measurement for data analysis (MacMullen & Denn, 2005). Interestingly, these individual fields have different research traditions with various rates and direction of information exchanges, and as a result, their citation patterns are quite different; therefore, the field-dependent patterns need to be characterized by analyzing their bibliometric data.

Standard bibliometric assessments of research performance in interdisciplinary fields pose challenges in making fair and comparable scientific publication evaluations for appointments of research staff and faculty promotions. Hirsch (2005) presented the *h*-index that was used to characterize the scientific performance of scientists (Cronin & Meho, 2006) and journals (Schubert & Glänzel, 2007). Wagner (2011) reviewed the approaches taken to assess interdisciplinary research. Different fields have their own citation patterns and have diverse impacts on interdisciplinary journals (Larivière & Gingras, 2010), and necessary field normalization has been proposed to evaluate citation performance (Zitt & Small, 2008).

Bibliometric analysis has been used in various analyses of the citation patterns of authors (Cronin & Overfelt, 1993; Lin, White, & Buzydlowski, 2002; Jeong, Lee, & Kim, 2009; Yan & Ding, 2010; Zhao & Strotmann, 2011), as well as of journal articles (Moed, 2005). Garfield's (1955) citation impact factor (IF) has been widely used in the assessment of journal quality. Impact factor normalization was proposed based on the ranking of the journals (Pudovkin & Garfield, 2004) or on the impact factor without adjustment for its value (Marshakova-Shaikevitch, 1996; Schubert & Braun, 1993; Schubert & Braun, 1996; Sen, 1992; Vinkler, 2002); however, it has been shown that only considering impact factor in this measurement can lead to biased interpretations of research (Moed, 2005). Field normalization methods, such as the mean normalized citation score (Moed, De Bruin, & Van Leeuwen, 1995) or citation rate (Schubert & Braun, 1986), were reported by comparing the average number of citations in a set of documents with the average number of citations in a given field (Glänzel et al., 2009). A journal-based normalization method, *J*-factor (Bell, Mittermaier, & Tunger, 2009), was proposed to make citation profile comparisons among different research institutes.

The bioinformatics field-dependent patterns were mixed with the multidisciplinary nature, making it even more difficult to delineate the emerging and complex interdisciplinary patterns through bibliometric study (Molatudi et al., 2009). Particularly, bibliometric analysis has been used for assessing the trends of bioinformatics research in particular countries (Guan & Gao, 2008; Molatudi et al., 2009), as well as the article key words, citation impact, and research collaboration (Patra & Mishra, 2006; Glänzel, Janssens, & Thisjs, 2009). However, in bioinformatics, there are two prominent factors in considering journal citation impact: article life span and citing propensity.

The article life span is represented by the obsolescence indicator such as cited-half-life (Avramescu, 1979; Wallace, 1086; Tsay, 1998; Egghe & Rousseau, 2 000a; Egghe & Rousseau, 2000b; Nicholas et al., 2005; Tona & Al, 2006; Larivière, Archambault, & Gingras, 2008). The normalization of citation frequency over the course of time is necessary (Kuo & Rupe 2007; Yu, Yang, & Liang, 2010), especially for journals in modern fields of science with short article life spans, such as biology. Articles in journals of more traditional fields of science, such as mathematics, tend to have a much higher longevity (Glänzel & Schoepflin, 1999).

The citing propensity reflected by the number of references for the article can be normalized to compute the citation impact (Biglu, 2008; Gomez-Sancho & Manceboon-Torrubia, 2009). Specifically, fractional counting (Small & Sweeney, 1995) was used for co-citation analysis and incorporated into journal normalization (Zitt & Small, 2008; Moed, 2010; Leydesdorff & Bornmann, 2011). Garfield (2002) reported that the number of references per article was higher in biological journals than in mathematics journals. Some bioinformatics journals are developed from fields like biology or mathematics, so it is reasonable to consider that these journals might inherit the journal aging patterns and citing propensity from the contributing subfields.

Instead of looking at the citation patterns of individual articles, or starting with impact factor without normalization for comparing journal performance, we analyzed the citation patterns across bioinformatics journals as well as their knowledge subfields. A normalization method was used to evaluate bioinformatics journal citation performance, taking into account the impacts of article life spans and citing propensity. Analyzing the citation patterns in bioinformatics journals could help bioinformatics researchers understand how the bioinformatics discipline evolves as a dynamic interdisciplinary field. The practicing information professionals or librarians could also utilize the bioinformatics citation patterns for understanding the users and building the collection.

The paper is organized as follows: 1) identification of the citation patterns (impact factor, cited-half-life, and references-per-article) in bioinformatics journals, 2) correlation and comparison of the citation patterns with their correspondent knowledge subfields, and 3) mathematical formulation to normalize these effects for journal impact analysis.

## Methods and data collection

### *Journal selection*

Glänzel et al. (2009) conducted a comparative analysis of publication activity and citation impact using a core of bioinformatics literature (25 journals). Since bioinformatics is a highly interdisciplinary field, journals that contribute to bioinformatics may be found in various disciplines. The selection of the bioinformatics journals in this study was from diverse sources, which included bioinformatics journals from the International Society of Computational Biology (http://www.iscb.org/iscb-publications-journals), Bioinformatics at France (http://www.bioinformatics.fr/journals.php), the Wikipedia bioinformatics journal list (http://en.wikipedia.org/wiki/List_of_bioinformatics_journals), PhamTao (http://www.pharmtao.com/Bioinfo/BioinformaticsJournals.htm), and journals under the "Mathematical and Computational Biology" section in the Web of Science (http://apps.isiknowledge.com) Science Journal Citation Report (SJCR). We comprised a comprehensive consensus list of 32 bioinformatics journals by eliminating those with fewer than 500 total citations reported in SJCR, those that have been published for less than 10 years, and those in review or newsletter formats. Individual journal citation data during the 2003– 2007 period, including IFs for two years and five years, cited-half-life, and references-per-article, were directly retrieved from the SJCR website in October 2009 .

[Insert Table 1 in here]

***Data statistical analysis***

SJCR categorized each bioinformatics journal into a discipline. In order to look at the relationships between the individual journals and their derived knowledge subfields, the aggregated citation data that represented their knowledge fields (disciplines), including impact factor and cited-half-life, were calculated. The Pearson correlation was analyzed to compare data from the impact factors, article life span, and citing propensity for each journal, as well as for their derived knowledge subfields.

***Bioinformatics journal impact normalization***

Citation impact adjustments were proposed to normalize the effects of differences in journal article life spans (cited-half-life) and propensity to cite (references-per-article), respectively. Previous research expressed R-impact factor (Kuo & Rupe, 2007) as $RIF_i(t) = IF_i(t) \times F_i^{-1}(1/2)$, where $IF_i(t)$ is impact factor for *ith* journal in year *t*, and

$F_i^{-1}(1/2)$ is cited-half-life (see the definitions in SJCR). R-impact factor (RIF) normalized the effect of article life span on the citation pattern. The impact of a journal's citing propensity can also be normalized by appending a new variable $R_i(t)$ to R-impact factor (Zitt & Small, 2008; Gomez-Sancho & Mancebon-Torrubia, 2009). The propensity adjusted impact factor (PRIF) was computed for each journal by the following equation:

$$PRIF_i(t) = IF_i(t) \times F_i^{-1}(1/2) \times R_i(t) \quad (1)$$

in (1) $R_i(t)$ is defined as:

$$R_i(t) = AVC(t) / AVC_i(t) \quad (2)$$

in (2) $AVC(t)$ was computed as:

$$AVC(t) = \sum_{i=1}^{N} \sum_{j=t_0}^{t} r_{ij} / \sum_{i=1}^{N} \sum_{j=t_0}^{t} p_{ij} \quad (3)$$

and $AVC_i(t)$ was computed as:

$$AVC_i(t) = \sum_{j=t_0}^{t} r_{ij} / \sum_{j=t_0}^{t} p_{ij} \quad (4)$$

$AVC(t)$ was defined as the average number of citations per article in a given list of journals, $AVC_i(t)$ was the average number of citations per article in a particular journal, $r_{ij}$ is total citation number in *jth* year for *ith* journal, $p_{ij}$ is citable article number for *ith* journal, and $N$ is the number of journals. Adjustments and a computation formula were proposed to normalize these two factors in citation impact. Journal rankings were compared before and after normalization.

## Results

### *Citation patterns in bioinformatics journals*

Table 1 showed 32 bioinformatics journals' impact factors displayed as IF(2): impact factor for two years, and IF(5): impact factor for five years, their cited-half-life scores, as well as their references-per-article scores based on seven disciplinary subcategories. As Table 1 indicates, some of the 32 journals have varied bibiliometric patterns in three variables of impact factor, cited-half-life, and references-per-article. IF(5) scores ranged from 1.06 (*Biometrical Journal:* Statistics) up to 9.67 (*Genomics*: Genetics). As for the cited-half-life, it was sorted from 2.4 years (*BMC Genomics*: Genetics) to over 10 years (e.g., *Mathematical Biosciences*: Math). The references-per-article scores ranked from 21 references-per-article (*Biometrika:* Statistics) to 55.4 references-per-article (*Genome Biology*: Genetics).

When comparing the low/high impact factors and cited-half-life, genetic and molecular biology journals (e.g., *Proteomics*, *Bioinformatics*, and *Genome Biology*) showed high impact factors and short life spans (Cunningham, 1995; McCain & Turner, 1989), while math and statistics journals (e.g., *Biometrika*, *Journal of Mathematical Biology*, and *Mathematical Bioscience*) demonstrated extremely long cited-half-life with low impact factors (Cunningham, 1995). To better analyze the relationships between journal life spans and impact factors of the journals, the disciplinary categories to which they belonged, as well as the IF(5) and cited-half-life scores were plotted in Figure 1. Mathematics and statistics bioinformatics journals showed longer article life spans and lower impact factors compared to other fields; bioinformatics journals related to genetics, biochemistry, and molecular biology demonstrated a sparse distribution for journal impact factors and article life spans. Figure 2 showed the relationship between the number of references per article and impact factor in bioinformatics journals. The bioinformatics journals with higher impact factors (e.g., genetics, molecular biology related) generally have more references per article. A Pearson correlation analysis of IF(2), IF(5), cited-half-life, and references-per-article indicated that IF(2) and IF(5) had a positive correlation ( $p$ <0.01) and that references-per-article had a positive ($p$ <0.01) correlation with IF(2) or with IF(5); however, cited-half-life did not have statistical correlation with any of the other factors.

[Insert Figure 1, 2 here]

### *Interrelationships with derived knowledge fields*

Further investigation was made of the relationship among the impact variables of the journals and their derived knowledge fields/subfields. For each subcategory's journals, the values for the whole subfield and the aggregated IF(2) and cited-half-life values for the particular subfield's group of bioinformatics journals were computed and summarized in Table 2. For example, within molecular biology category there were five bioinformatics-related journals, but within the whole subfield, 275 journals fell into the same category in the Science Journal Citation Report. Again, biology-related (molecular biology and biochemistry) bioinformatics journals had high impact factors with short life spans, while the math-related ones had low impact factors but long life spans (Table 2). The Pearson correlation analysis indicated that these bioinformatics journals' IF(2) and cited-half-life scores had positive correlations ( $p < 0.05$) with subfield's IF(2) and cited-half-life.

[Insert Table 2 in here]

### *Citation impact normalization*

Given the fact that bioinformatics journals' impact factors are field-dependent, bioinformatics journals with high impact factors are more likely to be found in rapidly expanding research areas (e.g. genetics, molecular biology), a short lived body of literature with more references per article (Seglen, 1997). If these patterns are sparse, normalizing them could be beneficial for comparison. For this reason, we considered impact normalization as a neutral estimate. The method used to normalize these field-dependent factors resulted in a single value that favored journals with longer cited-half-life and reduced the impact of the citing propensity effects. IF(2), IF(5), RIF(2), and PRIF(2) measurement indexes of the bioinformatics journals were computed, and their correspondent ranking orders were listed in Table 3. Bioinformatics research dealing with dynamic research subfields that have a high amount of activity and short article life spans, such as biochemistry and molecular biology, have a correspondingly high proportion of citations to recent publications—hence, higher journal impact factors—than, for example, mathematics. Particularly, the PRIF(2) ranking places math- and statistics-related journals (e.g., *Biometrics*, *Statistics in Medicine*) into higher ranks than the results derived from other methods, due to stable life spans of papers and low citing propensity for journals in these subfields (Table 3). Pearson correlation analysis indicated that the PRIF(2) had positive correlations ($p < 0.01$) with IF(2), IF(5), and RIF(2).

[Insert Table 3 in here].

## Discussion

Publication activity and citation habits can vary among subfields, making scientometric evaluation difficult (Glänzel & Schubert, 2003). Bioinformatics journals are solving particular biological problems in various subfields, and these journals inevitably inherit the citation patterns of other fields. Citation analysis in bioinformatics journals explored diverse citation characteristics and the interplay of relationships among impact factors, article life spans, and citing propensity. The journals published bioinformatics solutions related to genetic, biochemistry, and molecular biology are active and fast-developing areas while those related to mathematics and statistics are more stable with a slower development speed. Bioinformatics journals derived from the dynamic research subfields such as genetics and molecular biology accumulate a huge number of citations within a short time period, but the published articles quickly become obsolete (Marton, 1985).  Alternatively, articles from subfields like mathematics and statistics have a longer life span in the literature (Bensman, Smolinsky, & Pudovkin, 2010) but have a smaller portion of short term citations, leaving them with lower impact factors (Moed, Burger, Frankfort, & Van Raan, 1985). Subfields such as medical informatics depend greatly on the basic clinical research output and must translate the research findings from the bench work to the bedside; consequently, medical informatics journals eventually have fewer citations than the subfields of basic biological research (Seglen, 1997). Most of the bioinformatics journals are derived from the established journals in other disciplines and have inherited those subfields' bibliometric patterns.

The citation impact of a research field is directly proportional to the mean number of references per article, which varies considerably from field to field (Moed et al., 1985). In the dynamic research subfields (e.g., genetics, molecular biology), the high impact factor is related to the number of citable materials (Vinkler, 1996). A higher impact factor journal usually corresponds with a higher number of references per article (Zitt & Small, 2008; Biglu, 2008). Conversely, simply increasing the number of references for a journal publication would not necessarily "increase" the impact factor score. PRIF uncovered this "hidden bias" by normalizing it for journal citation comparisons. The positive correlations between impact factors with or without normalization indicated such adjustments of normalization were within the same ranges of the measurement with reasonable variations.

## Conclusion

Bioinformatics journals show scattered and field- dependant journal citation patterns. Article life spans reflect the pace of the evolving, dynamic natures of the subfields in bioinformatics. Those bioinformatics journals with higher impact factors generally come with higher citing propensity. Such a propensity simply increases the chances of citation in number but without providing real improvement for the journal citation impact. The citation impacts, as well as the article life spans of these journals, show similar patterns "inherited" from their correspondent knowledge fields. Based upon previous work and current study, a normalized impact factor PRIF was proposed to make a comparative analysis of the bioinformatics journals' citation impacts by normalizing the differences in impact factor, article life spans, and citing propensity.

This study also has its limits. The bibliometric data used for this research was limited to journal citation impacts and therefore this study focuses on journal level citation patterns. The future research will expend on the individual article life spans analysis conducted by Albarrán and Ruiz-Castillo (2011) and Walters (2011), as well as citation propensity in bioinformatics journals. Additionally, this study used the subfields of the journals within bioinformatics discipline as defined by the ISI Journal Citations Reports subject categories. Bioinformatics journals may belong to one, two, or even more of these subject categories. Only one category was chose as a representative subfield in this study. The results of statistics, however, showed diverse citation patterns of impact factor, article life spans, and citing propensity in bioinformatics journals, which can be well adjusted and normalized by the proposed approach.

## Acknowledgements

The authors would like to thank Stvilia Besiki and Krystal Bullers for their assistances, and the reviewers' useful comments, leading to a better final product.

## References

Albarrán, P., & Ruiz-Castillo, J. (2011). References made and citations received by scientific articles. *Journal of the American Society for Information Science and Technology*, 62(1), 40-49.

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

Ball, R., Mittermaier, B., & Tunger, D. (2009). Creation of journal-based publication profiles of scientific institutions – A methodology for the interdisciplinary comparison of scientific research based on the J-factor, *Scientometrics*, 81, No. 2 (2009) 381–392.

Bensman, S. J., Smolinsky, L. J., & Pudovkin, A. I. (2010). Mean citation rate per article in mathematics journals: Differences from the scientific model. *Journal of the American Society for Information Science and Technology, 61*(7), 1440–1463.

Biglu, M. (2008). The influence of references per paper in the SCI to impact factors and the Matthew Effect. *Scientometrics*, *74*(3), 453–470.

Brown, C.M.  (2003). The changing face of scientific discourse:  analysis of genomic and proteomic database usage and acceptance.  *Journal of the American Society for Information Science & Technology* 54(10): 926-938.

Cunningham, S. (1995). An empirical investigation of the obsolescence rate for information systems literature. (Working paper 95/16). Hamilton, New Zealand: University of Waikato, Department of Computer Science.

Cronin, B., & Overfelt, K. (1993). Citation-based auditing of academic performance. *Journal of the American Society for Information Science*, *45*(2), 61–72.

Cronin, B., & Meho, L. I. (2006). Using the *h*-index to rank influential information scientists. *Journal of the American Society for Information Science and Technology, 57*(9), 1275–1278.

Drmanac R. (2010). The Advent of Personal Genome Sequencing. *Genetic Medicine* 13, 188-190.

Egghe, L., & Rousseau, R. (2000a). The influence of publication delays on the observed aging distribution of scientific literature. *Journal of the American Society for Information Science*, *51*, 158–165.

Egghe, L., & Rousseau, R. (2000b). Aging, obsolescence, impact, growth, and utilization: Definitions and relations. *Journal of the American Society for Information Science*, *51*, 1004–1017.

Garfield, E. (2000), Use of Journal Citation Reports and Journal Performance Indicators in measuring short and long term journal impact. *Croatian Medical Journal*, *41*, 368–374.

Glänzel, W., Janssens, F., & Thisjs, B. (2009). A comparative analysis of publication activity and citation impact based on the core literature in bioinformatics, *Scientometrics*, *79*(1), 109-129.

Glänzel, W., & Schoepflin, U. (1999). A bibliometric study of reference literature in the sciences and social sciences. *Information Processing and Management*, *35*(3), 31-44.

Glänzel, W., & Schubert, A. (2003). A new classification scheme of science fields and subfields designed for scientometric valuation purposes. *Scientometrics*, *56*(3), 357–367.

Glänzel, W., Thijs, B., Schubert, A., & Debackere, K. (2009). Subfield-specific normalized relative parameters and a new generation of relational charts: Methodological foundations illustrated on the assessment of institutional research performance. *Scientometrics*, 78(1), 165–188.

Garfield, E. (1955). Citation indexes to science: A new dimension in documentation through association of ideas. *Science*, *122*, 108–111.

Guan, J., & Gao, X. (2008). Comparison and evaluation of Chinese research performance in the field of bioinformatics. *Scientometrics*, *75*(2), 357-379.

Gomez-Sancho, J., & Mancebon-Torrubia, M. (2009). The evaluation of scientific production: Towards a neutral impact factor. Scientometrics, 81(2), 435–458.

Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America, 102*(46), 16569–16572.

Jeong, S., Lee, S., & Kim, H. (2009). Are you an invited speaker? A bibliometric analysis of elite groups for scholarly events in bioinformatics. *Journal of the American Society for Information Science and Technology*, *60*(6), 1118–1131.

Kuo, W., & Rupe, J. (2007). R-impact: reliability-based citation impact factor. *IEE Transaction of Reliability*, *56*(3), 366–367.

Larivière, V., & Gingras, Y. (2010). On the relationship between interdisciplinarity and scientific impact. *Journal of the American Society for Information Science and Technology. 61*(1), 126-131.

Larivière,V., Archambault, É., & Gingras,Y. (2008). Long-term variations in the aging of scientific literature: From exponential growth to steady-state science (1900–2004). *Journal of the American Society for Information Science and Technology*, *59*(2), 288–296.

Leydesdorff, L., & Bornmann, L. (2011) How fractional counting affects the Impact Factor: Normalization in terms of differences in citation potentials among fields of science, Journal of the American Society for Information Science and Technology, 62(2), 217-229.

Lin, X., White, H.D., & Buzydlowski, J. (2003), Real-time author co-citation mapping for online searching. *Information Processing and Management*, 39 (5), 689-706.

MacMullen*,* W. *&* Denn*, S*. (2005), Information problems in molecular biology and bioinformatics. *Journal of the American Society for Information Science and Technology*, *56*(5), 447-456.

Marshakova-Shaikevich, I. (1996), The standard impact factor as an evaluation tool of science fields and scientific journals, *Scientometrics*, *35*(2), 283–290.

Marton J. (1985). Obsolescence or immediacy? Evidence supporting Price's hypothesis. *Scientometrics*, *7*, 145-153.

McCain, K., & Turner, K. (1989). Citation context analysis and aging patterns of journal articles in molecular genetics. *Scientometrics*, *17*(1-2), 127-163.

Moed, H.F., Burger, W., Frankfort, J., & Van Raan, A. (1985).The application of bibliometric indicators: important field and time dependent factors to be considered. *Scientometrics*, *8*,177-203.

Moed, H.F., De Bruin, R.E., & Van Leeuwen, T.N. (1995). New bibliometric tools for the assessment of national research performance: Database description, overview of indicators and first applications. *Scientometrics*, 33(3), 381–422.

Moed, H.F. (2005). Citation analysis of scientific journals and journal impact measures. *Current Science*, *89*, 1990–1996.

Moed, H. F. (2010). Measuring contextual citation impact of scientific journals. *Journal of Informetrics*, 4(3) 265-277.

Molatudi, M., Molotja N., & Pouris A. (2009). A bibliometric study of bioinformatics research in South Africa, *Scientometrics*, *81*, 47-59.

Nicholas, D., Huntington, P., Dobrowolski, T., Rowlands, I., Jamali M., H.R., & Polydoratou, P. (2005). Revisiting obsolescence and journal article decay. *Information Processing & Management*, *41*, 1441–1461.

Patra, S., & Mishra S. (2006). Bibliometric study of bioinformatics literature. *Scientometrics*, *67*(3), 477-489.

Pudovkin A. I., & Garfield E. (2004). Rank-normalized impact factor: a way to compare journal performance across subject categories. In: *Proceedings of the 67th Annual Meeting of the American Society for Information Science & Technology*. Vol 41. Silver Spring, Md: American Society for Information Science & Technology, pp. 507–515

Seglen P. (1997) Why the impact factor of journals should not be used for evaluating research. *BMJ*, *314*(7079), 498–502.

Schubert, A., & Braun, T. (1986). Relative indicators and relational charts for comparative assessment of publication output and citation impact. *Scientometrics, 9*(5), 281-291.

Schubert, A., & Braun, T. (1993). Reference standards for citation based assessments. *Scientometrics*, *26*(1), 21–35.

Schubert, A., & Braun, T. (1996), Cross-field normalization of scientometric indicators. *Scientometrics*, *36*(1), 311–324.

Schubert, A., & Glänzel, W. (2007). A systematic analysis of Hirsch-type indices for journals. *Journal of Informetrics*, 1(2), 179–184.

Sen, B.K. (1992). Documentation note normalized impact factor. *Journal of Documentation*, *48*(3), 318–325.

Small, H., & Sweeney, E. (1985). Clustering the Science Citation Index Using Co-Citations I. A Comparison of Methods,. *Scientometrics 7*, 391-409.

Tona, Y., & Al, U. (2006) Scatter and obsolescence of journals cited in theses and dissertations of librarianship. *Library and Information Science Research*, *28*, 281-296.

Tsay, M. (1998). Library journal use and citation half-life in medical science. *Journal of American Society of Information Science and Technology*, *49*(14), 1283–1292.

Vinkler, P. (1996). Relationships between the rate of scientific development and citations. The chance for a citedness model. *Scientometrics*, 35, 375-86.

Vinkler, P. (2002), Subfield problems in applying the Garfield (Impact) Factors in practice. *Scientometrics*, *53*(2), 267–279.

Wagner, C., Roessner, J., Bobb, K., Klein, J., Boyack, K., & Keyton, J., et al. (in press). Approaches to understanding and measuring Interdisciplinary Scientific Research (IDR): A review of the literature. *Journal of Informetrics*.

Wallace, D. (1986). The relationship between journal productivity and obsolescence. *Journal of the American Society for Information Science*, *37*(3), 136–145.

Walters, G. D. (2011). The citation life cycle of articles published in 13 American Psychological Association journals: A 25-year longitudinal analysis. *Journal of the American Society for Information Science and Technology*, 62(8), 1629-1636.

Yan, E., & Ding, Y. (2010). Weighted Citation: An indicator of an article's prestige. *Journal of the American Society for Information Science and Technology*, *61*(8), 1635-1634.

Yu, G., Yang, D., & Liang, W. (2010). Reliability-based citation impact factor and the manupulation of impact factor. *Scientometrics*, *83*(1), 259-270.

Zhao, D., & Strotmann, A. (2011). Counting first, last, or all authors in citation analysis: A comprehensive comparison in the highly collaborative Stem Cell research field. *Journal of the American Society for Information Science and Technology,* 62 (4): 654-676.

Zitt, M., & Small, H. (2008). Modifying the journal impact factor by fractional citation weighting: The audience factor. *Journal of the American Society for Information Science and Technology*, *59*, 1856–1860.

## Figures

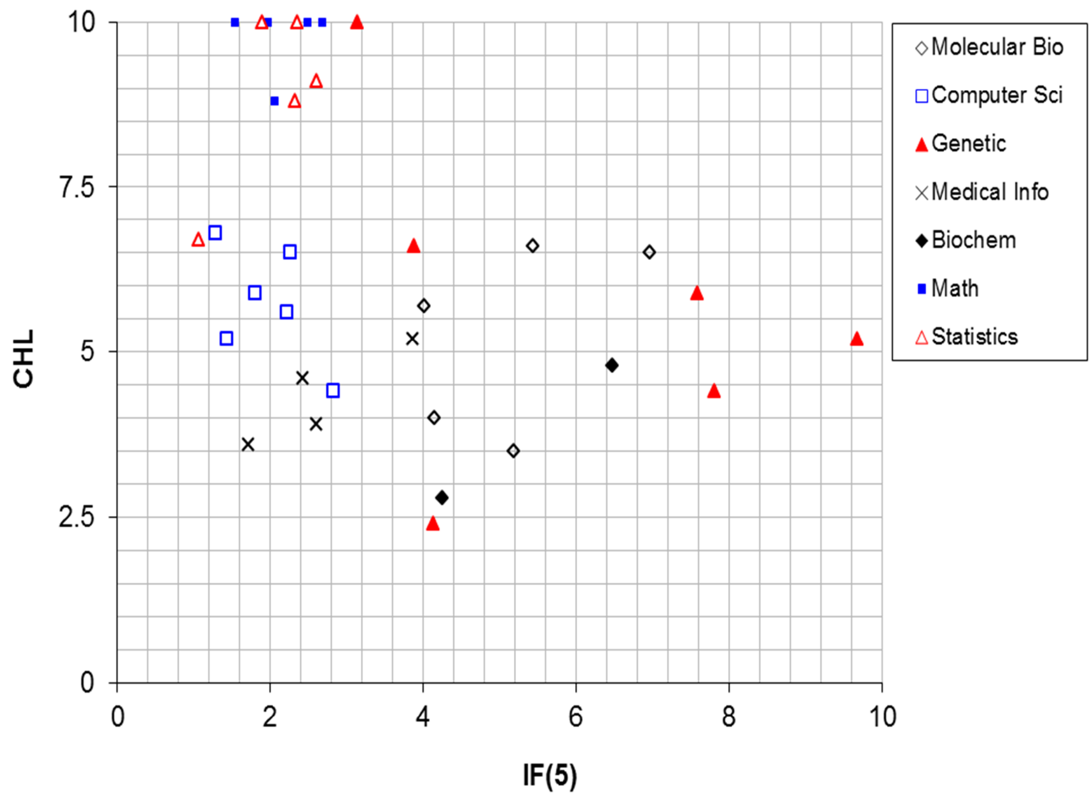


Figure 1. Scatter chart for bioinformatics journals' impact factor: IF(5) and Cited Half Life (CHL) grouped by JCR subject categories.

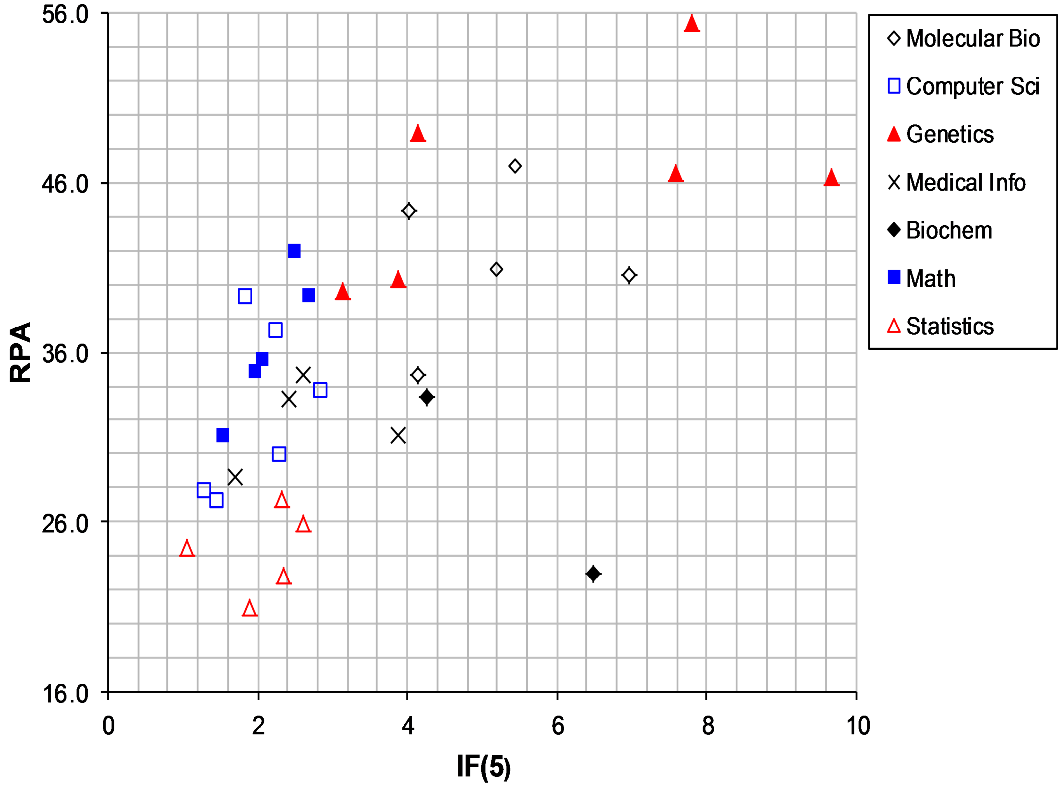


Figure 2. Scatter chart for bioinformatics journals' impact factor: IF(5) and References per article (RPA) group by JCR subject categories.

TABLE 1. Descriptive data of 32 Bioinformatics Journals

| **Subcategories** | **Bioinformatics Journal** | **Articles** | **Total Citations** | **IF(2)** | **IF(5)** | **Cited Half Life (2007)** | **Reference per article (2007)** |
|---|---|---|---|---|---|---|---|
| Biochemistry Research Method | Bioinformatics | 3166 | 20520 | 4.32 | 6.48 | 4.8 | 23.7 |
| | BMC Bioinformatics | 1734 | 7362 | 3.78 | 4.25 | 2.8 | 34.8 |
| Biochemistry& Molecular Biology | Metabolic Engineering | 204 | 718 | 3.52 | 4.14 | 4 | 39.9 |
| | Nucleic Acids Research | 5649 | 39362 | 6.88 | 6.97 | 6.5 | 40.3 |
| | Proteins-structure function and bioinformatics | 1919 | 7712 | 3.42 | 4.02 | 5.7 | 42.8 |
| | Proteomics | 2076 | 10777 | 4.59 | 5.19 | 3.5 | 42 |
| | Structure | 816 | 4442 | 5.4 | 5.44 | 6.6 | 45.1 |
| Computer Science Interdisciplinary | Artificial Intelligence in Medicine | 248 | 551 | 1.96 | 2.22 | 5.6 | 37.6 |
| | Computer Methods and Programs in Biomedicine | 503 | 646 | 1.22 | 1.28 | 6.8 | 24.5 |
| | Computers in Biology and Medicine | 419 | 598 | 1.27 | 1.43 | 5.2 | 26.8 |
| | IEEE Transactions on Information Technology in Biomedicine | 315 | 890 | 1.94 | 2.83 | 4.4 | 28.8 |
| | Journal of Computational Biology | 408 | 927 | 1.56 | 2.27 | 6.5 | 30.5 |
| | Journal of Molecular Modeling | 404 | 703 | 2.02 | 1.81 | 5.9 | 38.3 |
| Genetics & Heredity | BMC Genomics | 1131 | 4669 | 3.93 | 4.13 | 2.4 | 47 |
| | DNA Research | 153 | 593 | 3.61 | 3.88 | 6.6 | 41.9 |
| | Genome Biology | 927 | 7242 | 6.15 | 7.81 | 4.4 | 50.9 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | Genome Research | 1117 | 10810 | 10.2 | 9.67 | 5.2 | 46.7 |
| | Genomics | 857 | 2693 | 3.08 | 3.14 | 10 | 39 |
| | Human Molecular Genetics | 1745 | 13249 | 7.25 | 7.59 | 5.9 | 47.2 |
| Math, Computational Biology | Bulletin of Mathematical Biology | 414 | 849 | 1.74 | 2.05 | 8.8 | 35.5 |
| | Journal of Mathematical Biology | 310 | 611 | 1.58 | 1.97 | >10 | 30.2 |
| | Journal of Molecular Graphics & Modeling | 342 | 918 | 2.35 | 2.68 | >10 | 42.6 |
| | Journal of Theoretical Biology | 1613 | 4016 | 2.45 | 2.49 | 10 | 38 |
| | Mathematical Biosciences | 458 | 699 | 1.15 | 1.53 | >10 | 31.6 |
| Medical Informatics | International Journal of Medical Informatics | 454 | 1101 | 2.75 | 2.42 | 4.6 | 26.4 |
| | Journal of Biomedical Informatics | 256 | 667 | 1.92 | 2.61 | 3.9 | 32 |
| | Journal of the American Medical Informatics Association | 376 | 1461 | 3.43 | 3.87 | 5.2 | 31.6 |
| Statistics | Biometrical Journal | 345 | 365 | 1.11 | 1.06 | 6.7 | 22.9 |
| | Biometrics | 622 | 1463 | 1.97 | 2.35 | 10 | 22.4 |
| | Biometrika | 381 | 719 | 1.41 | 1.89 | 10 | 22.6 |
| | Statistical Methods in Medical Research | 145 | 377 | 2.18 | 2.6 | 9.1 | 35.9 |
| | Statistics in Medicine | 1343 | 3109 | 2.11 | 2.32 | 8.8 | 26.7 |

Note. IF(2)= Impact Factor for two years; IF(5)= Impact Factor for five years, data collected from 2003~2007.

TABLE 2. Bioinformatics journals and their subfields' Impact Factor and Cited Half Life.

| Subcategories | Bioinformatics Journal Name | IF(2) | Cited Half Life | Discipline IF(2) | Discipline Cited Half Life |
|---|---|---|---|---|---|
| Biochemistry Research Method | Bioinformatics<br>BMC Bioinformatics | 4.08 (2*) | 3.5 (2) | 3.27 (65) | 5.7 (65) |
| Biochemistry& Molecular Biology | Metabolic Engineering<br>Nucleic Acids Research<br>Proteins-structure function and bioinformatics<br>Proteomics<br>Structure | 3.44 (5) | 5.8 (5) | 4.24 (275) | 6.9 (275) |
| Computer Science Interdisciplinary | Artificial Intelligence in Medicine<br>Computer Methods and Programs in Biomedicine<br>Computers in Biology and Medicine<br>IEEE Transactions on Information Technology in Biomedicine<br>Journal of Computational Biology<br>Journal of Molecular Modeling | 1.6 (6) | 4.8 (6) | 1.55 (94) | 7 (94) |
| Genetics & Heredity | BMC Genomics<br>DNA Research<br>Genome Biology<br>Genome Research | 5.88 (6) | 4.7 (6) | 4.36 (138) | 6.2 (138) |

| | | | | | |
|---|---|---|---|---|---|
| | Genomics<br>Human Molecular Genetics | | | | |
| Math, Computational Biology | Bulletin of Mathematical Biology<br>Journal of Mathematical Biology<br>Journal of Molecular Graphics & Modeling<br>Journal of Theoretical Biology<br>Mathematical Biosciences | 2.07 (5) | 10 (5) | 2.84 (29) | 7 (29) |
| Medical Information | International Journal of Medical Informatics<br>Journal of Biomedical Informatics<br>Journal of the American Medical Informatics Association | 2.55 (3) | 4.8 (3) | 1.78 (20) | 6.8 (20) |
| Statistics | Biometrical Journal<br>Biometrics<br>Biometrika<br>Statistical Methods in Medical Research<br>Statistics in Medicine | 1.88 (5) | 10 (5) | 1.16 (92) | 10 (92) |

* Number of journals in subcategories or discipline.

TABLE 3: Bioinformatics journal normalized rankings using IF(2), IF(5), RIF(2), and PRIF(2)

| **PRIF(2) Rank** | **IF(2) Rank** | **Bioinformatics Journal Name** | **IF(2)** | **IF(5)** | **RIF(2)** | **PRIF(2)** |
|---|---|---|---|---|---|---|
| 1 | 1 | Genome Research | 10.18 | 9.67 | 52.94 | 34.57 |
| 2 | 3 | Nucleic Acids Research | 6.88 | 6.97 | 44.72 | 34.36 |
| 3 | 2 | Human Molecular Genetics | 7.25 | 7.59 | 42.78 | 28.25 |
| 4 | 7 | Bioinformatics | 4.32 | 6.48 | 20.74 | 27.46 |
| 5 | 21 | Biometrics | 1.96 | 2.22 | 19.7 | 27.21 |
| 6 | 14 | Genomics | 3.08 | 3.14 | 30.8 | 25.48 |
| 7 | 5 | Structure | 5.4 | 5.44 | 35.64 | 24.69 |
| 8 | 19 | Statistics in Medicine | 2.02 | 1.81 | 18.57 | 22.98 |
| 9 | 28 | Biometrika | 1.41 | 1.89 | 14.1 | 20.9 |
| 10 | 16 | Journal of Theoretical Biology | 2.45 | 2.49 | 24.5 | 19.95 |
| 11 | 18 | Statistical Methods in Medical Research | 2.18 | 2.6 | 19.84 | 19.51 |
| 12 | 10 | DNA Research | 3.61 | 3.88 | 23.83 | 17.82 |
| 13 | 17 | Journal of Molecular Graphics & Modeling | 2.35 | 2.68 | 23.5 | 17.6 |
| 14 | 12 | Journal of the American Medical Informatics Association | 3.43 | 3.87 | 17.84 | 17.31 |
| 15 | 4 | Genome Biology | 6.15 | 7.81 | 27.06 | 17.28 |
| 16 | 26 | Journal of Mathematical Biology | 1.58 | 1.97 | 15.8 | 16.14 |
| 17 | 15 | International Journal of Medical Informatics | 2.75 | 2.42 | 12.65 | 14.99 |
| 18 | 13 | Proteins-structure function and bioinformatics | 3.42 | 4.02 | 19.49 | 14.07 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 19 | 25 | Bulletin of Mathematical Biology | 1.67 | 1.71 | 15.31 | 13.59 |
| 20 | 6 | Proteomics | 4.59 | 5.19 | 16.07 | 12.13 |
| 21 | 31 | Mathematical Biosciences | 1.15 | 1.53 | 11.5 | 11.4 |
| 22 | 30 | Computer Methods and Programs in Biomedicine | 1.22 | 1.28 | 8.3 | 10.95 |
| 23 | 27 | Journal of Computational Biology | 1.56 | 2.27 | 10.14 | 10.91 |
| 24 | 11 | Metabolic Engineering | 3.52 | 4.14 | 14.08 | 10.62 |
| 25 | 32 | Biometrical Journal | 1.11 | 1.06 | 7.44 | 10.16 |
| 26 | 20 | Journal of Molecular Modeling | 1.97 | 2.35 | 11.92 | 9.89 |
| 27 | 9 | BMC Bioinformatics | 3.78 | 4.25 | 10.58 | 9.66 |
| 28 | 23 | IEEE Transactions on Information Technology in Biomedicine | 1.92 | 2.61 | 8.54 | 9.66 |
| 29 | 22 | Artificial Intelligence in Medicine | 1.94 | 2.83 | 10.98 | 9.09 |
| 30 | 29 | Computers in Biology and Medicine | 1.27 | 1.43 | 6.6 | 7.66 |
| 31 | 24 | Journal of Biomedical Informatics | 1.74 | 2.05 | 7.49 | 6.7 |
| 32 | 8 | BMC Genomics | 3.93 | 4.13 | 9.43 | 6.41 |

* IF(2): Impact factor for two years, IF(5): Impact factor for five years, RIF(2): R-impact factor for two years, PRIF(2): Propensity adjusted impact factor for two years