On the Relationship between Author Collaboration and Impact of Scientific Publications

Rodrigo Costas and Cornelis van Bochove

rcostas@cwts.leidenuniv.nl; CBochove@cwts.leidenuniv.nl
Center for Science and Technology Studies (CWTS). Leiden University.

Introduction
Scientific collaboration has been widely reported to have positive effects on the impact of scientific publications. From a bibliometric point of view, publications in collaboration have been linked to higher impact (Katz & Martin, 1997; Persson & al., 2004). There is a growing interest in the study of collaboration in its different forms, especially regarding its potential benefits to scientific activity and research performance (Lee & Bozeman, 2005). This study presents a general analysis of the relationship of the impact of publications with the number of authors that appear in the by-line of papers. The main idea is to explore how this relationship behaves and to explore if there is any model to describe this relationship. Although this is a relatively simple analysis, to the best of our knowledge it has not been carried out before from a global perspective.

Objectives
The main objective thus is to study the relationship between impact and the number of authors of papers from a global perspective, both in the aggregate and by fields of science. A second objective is to propose a model that can explain this relationship and to analyze the fit of the model in different scientific disciplines.

Methodology
The data used in this study are all the publications covered by the Web of Science during the period 1981-2005 in all scientific fields. Citations have been calculated for all the publications considered, using a fixed citation window of 6 years. In addition, an analysis based on several major disciplines is also performed.

Several impact indicators have been calculated:
- CPP: ration of citations per publication.
- MNCS (Mean Normalized Citation Score). This is the field-normalized indicator of impact (see Waltman & al., 2011).

Throughout the analysis the natural logarithm of the impact indicators has been used in the presentation of the figures and models.

Results
Analysis of the total set of publications
Figure 1 shows the relationship between the mean impact of the publications and the level of co-authorship. In this case indicators including (+sc)/excluding self-citations (-sc) are considered.

Figure 1. Relationship between number of authors and impact (CPP & MNCS)—Total set of publications.

The relation between the impact of publications and the number of authors involved in them is positive and increasing, thus supporting the statement that publications with more authors (more collaboration) also have more impact. The differences between the figures for the two indicators are minor; therefore in the analysis per discipline we use the non-standardized CPP indicator only.
It is remarkable that the increment in impact is not linear but logarithmic (statistically, the logarithmic model outperforms other models like the polynomial, linear, etc.). In all the cases $R^2$ is higher than 0.80, thus suggesting that the model fits the distribution very well.

Inspection of the indicators with and without self-citations shows that the relationship between the two lines is parallel. Thus we can assume that the pattern is not affected by self-citations. In both cases (CPP and MNCS) the growth in impact per collaboration level is lower when self-citations are excluded, reinforcing the idea that self-citations happen more frequently in collaborative papers (Costas et al, 2010). In view of this result we use only data including self-citations in our analysis per fields.

Selection of some fields
Figure 2 presents the analysis and the fit of the logarithmic regression for 3 different disciplines.

**Medical & Life Sciences**
\[ y = 0.7429 \ln(x) + 1.5355 \quad R^2 = 0.9431 \]

**Law, Arts & Humanities**
\[ y = 0.7493 \ln(x) + 0.3422 \quad R^2 = 0.895 \]

**Multidisciplinary Journals**
\[ y = 1.0982 \ln(x) + 2.0321 \quad R^2 = 0.9721 \]

**Figure 2.** Relationship between author collaboration and impact (CPP)—3 disciplines.

Figure 2 confirms the good fit of the logarithmic regression in major disciplines ($R^2$ values normally higher than 0.80). The relationship between author collaboration and impact for publications in those journals classified as multidisciplinary shows the highest fit of the logarithmic regression ($R^2=0.972$) as well as the highest growth in impact with collaboration (2.03). However, even the discipline ‘Law, Arts & Humanities’ fits the logarithmic regression ($R^2=0.895$) very well and displays a clear growth in impact with collaboration.

**Discussion & Conclusions**
Although the analysis of scientific collaboration has been a recurrent topic in bibliometrics and research policy during the last decades, to the best of our knowledge there are no studies that analyze the pattern of relationship between the level of authorship of publications and their impact from a global perspective and provide a model that describes this relationship as the present paper does.

The main result of this study is that a pattern of regularity in scientific production exists. This pattern is that the relationship between author collaboration and the impact of the publications does not follow a linear relationship but a logarithmic one. The pattern is observed both with and without field normalization of the impact indicators, and both with and without self-citations. The logarithmic pattern exists in the major disciplines, including the medical sciences and the humanities and social sciences.

The patterns found here can have science policy relevance since they suggest that measures, such as those by the EU, to stimulate (international) collaboration, also lead to a higher scientific impact. However, further research is needed to corroborate whether collaboration also increases the impact per person-time spent on the publications.

**References**