# Statistical Analysis of the Effect of Equations on Citations

Cited in the paper:

**Equation-dense papers receive fewer citations—in physics as well as biology**

New Journal of Physics

A. D. Higginson, T. W. Fawcett

Centre for Research in Animal Behaviour, College of Life and Environmental Sciences, University of Exeter, Exeter EX4 4QG, UK.

adhigginson@gmail.com, tim.fawcett@cantab.net

Kollmer et al [1] presented data on the number of equations and citations for the set of papers published in volumes 94 and 104 of *Physical Review Letters.* This is a convenient data set because none of the papers have appendices, so all equations must be present in the main text of the article, and almost all are four pages long, so Kollmer et al [1] could simply divide the number of equations by four to get an approximate measure of equation density (Thorsten Pöschel, personal communication). Equation density ranged from 0 to 8.75 equations per page (mean ± SEM: 1.237 ± 0.034). The number of citations varied widely, ranging from 0 to 809 (mean ± SEM: 30.629 ± 1.026). As for the data in our original paper, the physics data were extremely over-dispersed (the variance-to-mean ratio was in excess of 65), which results from the tendency for citations to attract ever more citations to a paper. In this data set ($n$ = 1906) the clustering is even more extreme than in the biology data (estimated clumping parameter $k$ = 0.475; [2]). We therefore again used a negative binomial model [3], specified by the function glm.nb in the MASS library in *R* [4], which takes into account the degree to which the data cluster together [5]. We modelled variation in the number of citations (dependent variable) as a function of equation density, journal volume and the interaction between these explanatory variables. This analysis showed that equation density has a statistically significant negative effect on the number of citations, leading on average to 6% fewer citations for each additional equation per page (Table 1, *all papers*).

To allay the concerns expressed by Kollmer et al [1] about heavily cited papers possibly affecting the result, we omitted papers with over 100 citations (Table 1, *not heavily cited*). This results in an even stronger negative effect of equation density (8% fewer citations per equation per page). Finally, to check that the effects were not merely due to papers containing some equations being generally less well cited than those containing none, we omitted papers containing zero equations. The negative effect of equation density was weaker (5% fewer citations for each additional equation per page), but still statistically significant (Table 1, *equation-containing papers*).


## ACKNOWLEDGEMENTS

This work was supported by a European Research Council Advanced Grant (250209) awarded to Alasdair Houston and NERC Independent Research Fellowship (NE/L011921/1) awarded to A. D. H. We are indebted to the American Physical Society and *Physical Review Letters* for granting access to their article data and Jonathan Kollmer, Thorsten Pöschel and Jason Gallas for supplying it.


## REFERENCES


[1] Kollmer JE, Pöschel T, Gallas JAC 2015 *New J Physics* 17 013036.

[2] Crawley MJ 2007 *The R Book* John Wiley & Sons, Chichester, UK.

[3] White GC, Bennetts RE 1996 *Ecology* 77 2549

[4] R Core Team 2014. *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org

[5] Bolker BM 2008 *Ecological Models and Data in R* Princeton Univ Press, Princeton, NJ.

**Table 1. Variables affecting the number of citations for (*A*) all papers, (*B*) papers with fewer than 100 citations, (*C*) papers containing at least one equation and with fewer than 100 citations.**

| | (*A*) all papers $n = 1906$ | | | (*B*) not heavily cited $n = 1809$ | | | (*C*) equation-containing papers $n = 1346$ | | |
|---|---|---|---|---|---|---|---|---|---|
| parameter | OR (95% CI) | Wald *z* | *P* | OR (95% CI) | Wald *z* | *P* | OR (95% CI) | Wald *z* | *P* |
| intercept | 17.86 (16.53–19.31) | 72.85 | < 0.001 | 17.16 (16.01–18.39) | 80.67 | < 0.001 | 15.68 (14.28–17.21) | 57.97 | < 0.001 |
| equation density | **0.94** (0.90–0.98) | −3.09 | 0.002 | **0.92** (0.88–0.95) | −4.38 | < 0.001 | **0.95** (0.91–0.99) | −2.41 | 0.016 |
| journal volume | **2.72** (2.44–3.03) | 18.10 | < 0.001 | **1.92** (1.74–2.12) | 12.98 | < 0.001 | **1.98** (1.74–2.26) | 10.11 | < 0.001 |
| equation density × volume | 0.99 (0.93–1.05) | –0.37 | 0.709 | 1.05 (1.00–1.11) | 1.88 | 0.061 | 1.04 (0.98–1.10) | 1.25 | 0.210 |

The table shows statistical results from a generalized linear model with a negative binomial error structure. For a unit increase in the explanatory variable, the number of citations changes by a factor given by the odds ratio (OR), shown here with a 95% confidence interval (CI). For example, an OR of 0.94 implies a decrease of 6 per cent, while an OR of 1.05 implies an increase of 5 per cent. Significant effects ($P < 0.05$) based on the Wald *z* statistic are highlighted in **bold**.