PLOS ONE

Publication cycle: A study of the Public Library of Science (PLoS)

--Manuscript Draft--

Manuscript Number:	PONE-D-15-29777
Article Type:	Research Article
Full Title:	Publication cycle: A study of the Public Library of Science (PLoS)
Short Title:	Publication cycle: A study of the Public Library of Science (PLOS)
Corresponding Author:	Chris H.J. Hartgerink, MSc. Tilburg University Tilburg, NETHERLANDS
Keywords:	publishing; peer-review; plos; metadata; poisson
Abstract:	Publications are the driving force in current age academia. However, publishing is a tedious process and can take a considerable amount of time. Previous research has barely investigated whether parts of the publication cycle (i.e., review and production process) can be predicted based on metadata available for all research papers. The predictive value of metadata was investigated in this study with three predictors: (i) the number of authors, (ii) the length of the manuscript, and (iii) the presence of competing interests. Additionally, these models inspect changes in the publication cycle throughout the years. Model results indicate that the review and production times cannot be predicted by the included metadata of research papers. Results also indicate review times have doubled throughout the last decade for PLoS journals, which are currently estimated between 150-250 days on average. Production times, however, have remained highly stable throughout the last decade around an estimated mean 50 days. The results of these analyses indicate that review- and production times cannot be predicted by metadata, given a certain year-specific mean.
Order of Authors:	Chris H.J. Hartgerink, MSc.
Opposed Reviewers:	
Additional Information:	
Question	**Response**
Financial Disclosure Please describe all sources of funding that have supported your work. A complete funding statement should do the following: Include **grant numbers and the URLs** of any funder's website. Use the full name, not acronyms, of funding institutions, and use initials to identify authors who received the funding. **Describe the role** of any sponsors or funders in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. If they had no role in any of the above, include this sentence at the end of your statement: "*The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*" If the study was **unfunded**, provide a	The author received no specific funding for this work.

statement that clearly indicates this, for example: "*The author(s) received no specific funding for this work*." * typeset	
Competing Interests You are responsible for recognizing and disclosing on behalf of all authors any competing interest that could be perceived to bias their work, acknowledging all financial support and any other relevant financial or non-financial competing interests. Do any authors of this manuscript have competing interests (as described in the PLOS Policy on Declaration and Evaluation of Competing Interests)? **If yes**, please provide details about any and all competing interests in the box below. Your response should begin with this statement: *I have read the journal's policy and the authors of this manuscript have the following competing interests:* **If no** authors have any competing interests to declare, please enter this statement in the box: "*The authors have declared that no competing interests exist.*" * typeset	The author has declared that no competing interests exist.
Ethics Statement You must provide an ethics statement if your study involved human participants, specimens or tissue samples, or vertebrate animals, embryos or tissues. All information entered here should **also be included in the Methods section** of your manuscript. Please write "N/A" if your study does not require an ethics statement. **Human Subject Research (involved human participants and/or tissue)**	N/A

All research involving human participants must have been approved by the authors' Institutional Review Board (IRB) or an equivalent committee, and all clinical investigation must have been conducted according to the principles expressed in the Declaration of Helsinki. Informed consent, written or oral, should also have been obtained from the participants. If no consent was given, the reason must be explained (e.g. the data were analyzed anonymously) and reported. The form of consent (written/oral), or reason for lack of consent, should be indicated in the Methods section of your manuscript.	
Please enter the name of the IRB or Ethics Committee that approved this study in the space below. Include the approval number and/or a statement indicating approval of this research.	
Animal Research (involved vertebrate animals, embryos or tissues) All animal work must have been conducted according to relevant national and international guidelines. If your study involved non-human primates, you must provide details regarding animal welfare and steps taken to ameliorate suffering; this is in accordance with the recommendations of the Weatherall report, "The use of non-human primates in research." The relevant guidelines followed and the committee that approved the study should be identified in the ethics statement.	
If anesthesia, euthanasia or any kind of animal sacrifice is part of the study, please include briefly in your statement which substances and/or methods were applied.	
Please enter the name of your Institutional Animal Care and Use Committee (IACUC) or other relevant ethics board, and indicate whether they approved this research or granted a formal waiver of ethical approval. Also include an approval number if one was obtained.	

Field Permit Please indicate the name of the institution or the relevant body that granted permission.	
Data Availability PLOS journals require authors to make all data underlying the findings described in their manuscript fully available, without restriction and from the time of publication, with only rare exceptions to address legal and ethical concerns (see the PLOS Data Policy and FAQ for further details). When submitting a manuscript, authors must provide a Data Availability Statement that describes where the data underlying their manuscript can be found. Your answers to the following constitute your statement about data availability and will be included with the article in the event of publication. **Please note that simply stating 'data available on request from the author' is not acceptable.** ***If, however, your data are only available upon request from the author(s), you must answer "No" to the first question below, and explain your exceptional situation in the text box provided.*** Do the authors confirm that all data underlying the findings described in their manuscript are fully available without restriction?	Yes - all data are fully available without restriction
Please describe where your data may be found, writing in full sentences. **Your answers should be entered into the box below and will be published in the form you provide them, if your manuscript is accepted.** If you are copying our sample text below, please ensure you replace any instances of XXX with the appropriate details. If your data are all contained within the paper and/or Supporting Information files, please state this in your answer below. For example, "All relevant data are within the paper and its Supporting Information files." If your data are held or will be held in a public repository, include URLs, accession numbers or DOIs. For example, "All XXX files are available from the XXX database (accession number(s) XXX, XXX)." If this information will only be available after acceptance, please indicate this by ticking the box below.	Data and analysis scripts are available in version-controlled format at the Open Science Framework at https://osf.io/53sn9/ without restrictions.

If neither of these applies but you are able to provide details of access elsewhere, with or without limitations, please do so in the box below. For example: “Data are available from the XXX Institutional Data Access / Ethics Committee for researchers who meet the criteria for access to confidential data.” “Data are from the XXX study whose authors may be contacted at XXX.” * typeset	
Additional data availability information:	

July 7, 2015

Dear PLOS staff,

I hereby submit my manuscript ‘Publication cycle: A study of the Public Library of Science (PLOS)’ I would appreciate it if you could consider my work for publication in *PLoS ONE*. This is an original manuscript, and is not under consideration elsewhere. The main text of the manuscript is 1,556 words long and is accompanied by 2 figures, 2 tables, and 2 supplementary files.

This paper is the first, to my knowledge, to move beyond describing data on publication times across journals. More specifically, I model the time it takes for a submitted manuscript to be accepted and published based on metadata available for all research papers published in PLoS journals. The results indicate article metadata does not systematically predict the length of the review process, except for the year the paper was published in. Besides providing insight into the publication cycle, the article is also a new application of the data available in the PLoS API.

Please note I made all research files available on the Open Science Framework (OSF). The link to this OSF page is provided in the manuscript. Hence, it is likely reviewers will find out I am the author. Personally, I do not consider this problematic.

I look forward to your reply and hope you will find my study of publication time across PLoS journals intriguing for review.

Kind regards,

Chris H.J. Hartgerink

Tilburg University
Warandelaan 2, 5037AB, Tilburg
c.h.j.hartgerink@tilburguniversity.edu

Publication cycle: A study of the Public Library of Science (PLOS)

C.H.J. Hartgerink[1]

1 Department of Methodology and Statistics, Tilburg University, Tilburg, the Netherlands

*Corresponding author

E-mail: c.h.j.hartgerink@tilburguniversity.edu

Abstract

Publications are the driving force in current age academia. However, publishing is a tedious process and can take a considerable amount of time. Previous research has barely investigated whether parts of the publication cycle (i.e., review and production process) can be predicted based on metadata available for all research papers. The predictive value of metadata was investigated in this study with three predictors: (i) the number of authors, (ii) the length of the manuscript, and (iii) the presence of competing interests. Additionally, these models inspect changes in the publication cycle throughout the years. Model results indicate that the review and production times cannot be predicted by the included metadata of research papers. Results also indicate review times have doubled throughout the last decade for PLoS journals, which are currently estimated between 150-250 days on average. Production times, however, have remained highly stable throughout the last decade around an estimated mean 50 days. The results of these analyses indicate that review- and production times cannot be predicted by metadata, given a certain year-specific mean.

Keywords: publishing, peer-review, plos, metadata

Science communication is primarily based on publishing research results in research papers. Anecdotally, authors feel that the publication cycle takes too long [1]. A better understanding of the publication lag could provide solace when feelings of substantial delay occur, where the main question is whether there are predictive factors of time taken from submission to publication. This paper tries to model publication times for the Public Libary of Science (PLoS) journals with metadata available for resesarch papers. The PLoS journals include PLoS Medicine, PLoS Biology, PLoS ONE, PLoS Pathogens, PLoS Genetics, PLoS Computational Biology, PLoS Neglected Tropical Diseases, and PLoS Clinical Trials (which was later merged into PLoS Medicine).

Previous research indicated that statistically nonsignificant results take longer to be published [2], review times have decreased [3], and that the amount of figures or tables does not predict publication time [4]. Other research into the academic publication cycle has focused on rejection rates of submitted manuscripts or the types of decisions made after the peer-review process [5]. These studies primarily relied on sampling research papers from journals, but with the rise of APIs and scrapers to mine the literature [6] such sampling is becoming redundant. In this paper, I analyze the entire population of PLoS research articles and split between predicting review time (i.e., time from submission through acceptance) and production time (i.e., time from acceptance through publication) in order to investigate whether publication time can be predicted with paper metadata.

Method

Article level data was collected from all PLoS journal research papers with v0.5 of the `rplos` package [7] in `R` v3.2.0 [8]. The dataset was collected on July 4, 2015 and is available via

https://osf.io/53sn9/. Research papers without the following were excluded: journal name, publication dates (i.e., submitted, accepted, and published), and problematic publication dates. Problematic publication dates include being published before accepted, accepted before submitted, or accepted at the same time as submitted.

The full publication cycle was split into the review process and the production process. The full publication cycle is the number of days between submission and publication, whereas the review process is the number of days between submission and acceptance; the production process is the number of days between acceptance and publication. The number of days for each element of the publication cycle was modeled with a Poisson regression model. A Poisson regression model is a linear regression model for count variables and assumes equal mean and variance (i.e., dispersion = 1). The data showed overdispersion (i.e., dispersion > 1) and quasi-likelihood estimation was used to correct for the violated dispersion assumption.

Model predictors were year of publication, presence of competing interests, number of pages, and number of authors. The reasoning behind these predictors was as follows. Competing interests could increase publication time when disputed by editors and authors are subsequently asked to explain. Number of pages could increase publication time due to longer reviews in both time taken to complete review, the length of the review, and increased production efforts required. Number of authors could influence the time it takes for authors to reach consensus on the response letter and potential other edits during the publication process. Squared predictors were included for number of pages and number of authors due to non-linear relations in scatterplots with review- and production days. Additionally, the number of authors and the number of pages were mean centred to provide meaningful intercept estimates.

Considering that the data are the population of data for PLoS research papers, statistical inference testing is not applied. Moreover, note that PLoS Clinical Trials was merged into PLoS Medicine in 2007 and only started in 2006, which is why other years are not included in estimates for this journal.

Results

Descriptive results

The collected dataset includes information on 140,674 research papers. Across all journals, the median publication cycle is 152 days, with the majority of this being the review process (i.e., median 111 days) and not the production process (i.e., median 38 days). Table **Error! Reference source not found.** specifies these numbers per journal and indicates PLoS ONE has the fastest review process, whereas PLoS Medicine has the longest review process (median difference = 69). PLoS Clinical Trials had the longest production process, compared to PLoS ONE (median difference = 16). S1 Figure includes plots of observed median review- and production times per journal.

Table 1. Descriptive statistics per journal, with publication-, review-, and production time in median.

	# Articles	Publication time	Review time	Production time
ONE	122,398	147	107	36
Clinical Trials	44	180.5	125	52
Genetics	4,741	182	131	50
Neglected Tropical Diseases	2,999	183	133	45
Pathogens	3,992	183	139.5	43
Biology	2,015	190	141	46
Computational Biology	3,423	199	148	48
Medicine	1,062	230.5	176	47
Overall	140,674	152	111	38

These differences in the review- and production speed could be a consequence of increased efficiency or stricter publication criteria. PLoS ONE contains 122,398 papers and is considered a megajournal (i.e., not field specific or selective in topic). On the other hand, the other journals are more similar to traditional journals in their criteria for publication (e.g., originality of research). PLoS Medicine, for example, contains 'only' 1,062 papers, indicating a large disparity with PLoS ONE.

Correlations indicate that the total publication cycle is almost perfectly correlated with review time ($\rho = .976$). This indicates that 95% of the variance in publication cycle is explained by the review time and that the production process seems an additive random process that is not predicted by the time taken to get a paper accepted.

Aggregate model results

Poisson model estimates for all journals together indicate that both review- and production time are only predicted by year. Coefficients in Table 2 indicate negligible predictive effects of number of authors, number of pages, and presence of competing interests (i.e., $b \leq |.017|$). Dummy coefficients indicate that review time has increased, whereas production time has fluctuated around 50 days. Besides the effect of year, the results indicate review time is a random process.

Table 2: Table 2. Poisson regression model estimates for review- and production time.

	Estimate (review)	Estimate (production)
Intercept	4.18370	4.24677
Authors (centred)	0.00176	0.00582
Authors^2 (centred)	-0.00001	-0.00001
Pages (centred)	-0.00084	0.00012
Pages^2 (centred)	-0.00010	-0.00011
Conflict of interest	-0.01713	0.00551
2004	0.68758	0.10155

2005	0.74031	-0.12891
2006	0.69579	-0.10830
2007	0.55104	-0.45996
2008	0.62225	-0.56911
2009	0.59525	-0.56514
2010	0.66045	-0.66266
2011	0.65463	-0.56665
2012	0.73687	-0.47161
2013	0.73887	-0.36991
2014	0.77532	-0.53661
2015	0.84229	-0.29643

The estimated mean review- and production time are depicted in Figure 1. For review time, the estimates are increasing in a non-linear fashion, with a short decreasing trend 2006 and 2008. The estimated mean review time has climbed to approximately 150 days since 2003. Estimated mean production time fluctuates around 50 days. The journal specific model results are described next.

Fig. 1. Mean estimated review- (top) and production (bottom) time in days across all PLoS journals, including loess curves.

Journal model results

When the results are specified per journal, model estimates are similar to the aggregate results described previously. Most journal specific models included no meaningful effect for number of authors, number of pages, or presence of competing interests on either the review- or production time. Only for PLoS Clinical Trials and PLoS Biology the presence of competing interests had a noteworthy effect on review- and production time (b = .112 and b = .106, respectively). This indicates that competing interests increase review- and production time by a factor of approximately 1.1 for Clinical Trials and PLoS Biology. All individual coefficients

per journal for both review- and production time are available in S2 File. Figure 2 plots the mean estimated review- and production times for each journal.

Fig. 2. Mean estimated review- (top) and production (bottom) time in days per PLoS journals, including loess curves (top) and regression lines (bottom).

Substantial variability is observed in estimated mean review times across journals, but all journals show an increasing time taken to complete the review process. In accordance with the descriptive statistics given earlier, PLoS Medicine has the longest estimated mean review time, whereas PLoS ONE is the fastest. As of 2015, the review process takes between 150-250 days on average and is less variable across journals than in the preceding years.

The estimated mean production times are highly consistent across journals and show less fluctuation than the aggregate results. The estimated mean review time is approximately 50 days across journals, across years.

Discussion

The results of this population level investigation of the PLoS publication cycle indicates that review times have doubled to 150-250 days in the last decade, production time has remained relatively stable at 50 days, and that the publication cycle is not substantially predicted by article metadata. The lack of predictive value of length of a manuscript, number of authors, or the presence of competing interests indicates that the publication cycle might be more a random- than a structured process.

It is noteworthy that, with the development of new editorial systems, the production times for research papers have remained stable in the last decade. Only recently, as of January 1

2015, PLoS has introduced a new set of manuscript guidelines to improve automatization of the production process. Note that the results in this paper show no systematic effect of this, or any previous, adjustment to the production process. The current system might provide this effect in the (near) future, but has not yet.

The increase in review time is substantial and begs the question why this review time has doubled. The increase in review times could be due to any amount of factors, ranging from increased difficulty of finding reviewers through authors taking longer to reply to reviewer comments. That review times are not predicted by the included metadata, however, eliminates these properties of papers as explanatory factors for increased review times. If, for example, the length of the manuscript increased throughout the decade and this explained the increased review time, the effect of year would disappear after controlling for manuscript length. This clearly was not the case.

In sum, authors are left guessing how long it takes for their paper to be published, where this paper indicates that the duration of the publication cycle might be random in some sense. More specifically, publication time seems to only be subject to trends throughout the years and not paper specific characteristics. The trends in the number of review days seem particularly strong, where the doubling of the review time is concerning.

References

[1] Himmelstein D (2015). Publication delays at PLOS and 3,475 other journals. URL http://web.archive.org/web/20150701055009/http://blog.dhimmel.com/plos-and-publishing-delays/. Accessed: 2015-7-1.

[2] Ioannidis, JPA (1998) Effect of the statistical significance of results on the time to completion and publication of randomized efficacy trials. JAMA 279: 281-286.

[3] Lyman RL (2013) A three-decade history of the duration of peer review. Journal of Scholarly Publishing 44: 211-220.

[4] Lee DT, LaCombe J, Chung CK, Ahlia K, Lee GK (2013) Lag-time to publication in plastic surgery potential impact on the timely practice of evidence-based medicine. Annals of Plastic Surgery 71: 410-414.

[5] Rosenkrantz AB, Harisinghani M (2015) Metrics for original research articles in the AJR: From first submission to final publication. American Journal of Roentgenology 204: 1152–1156.

[6] Smith-Unna R, Murray-Rust P (2014) The ContentMine scraping stack: Literature-scale content mining with community-maintained collections of declarative scrapers. D-Lib Magazine 20.

[7] Chamberlain S, Boettiger C, Ram K (2015) rplos: Interface to the search API for the 'PLoS' Journals. URL http://CRAN.R-project.org/package=rplos. R package version 0.5.

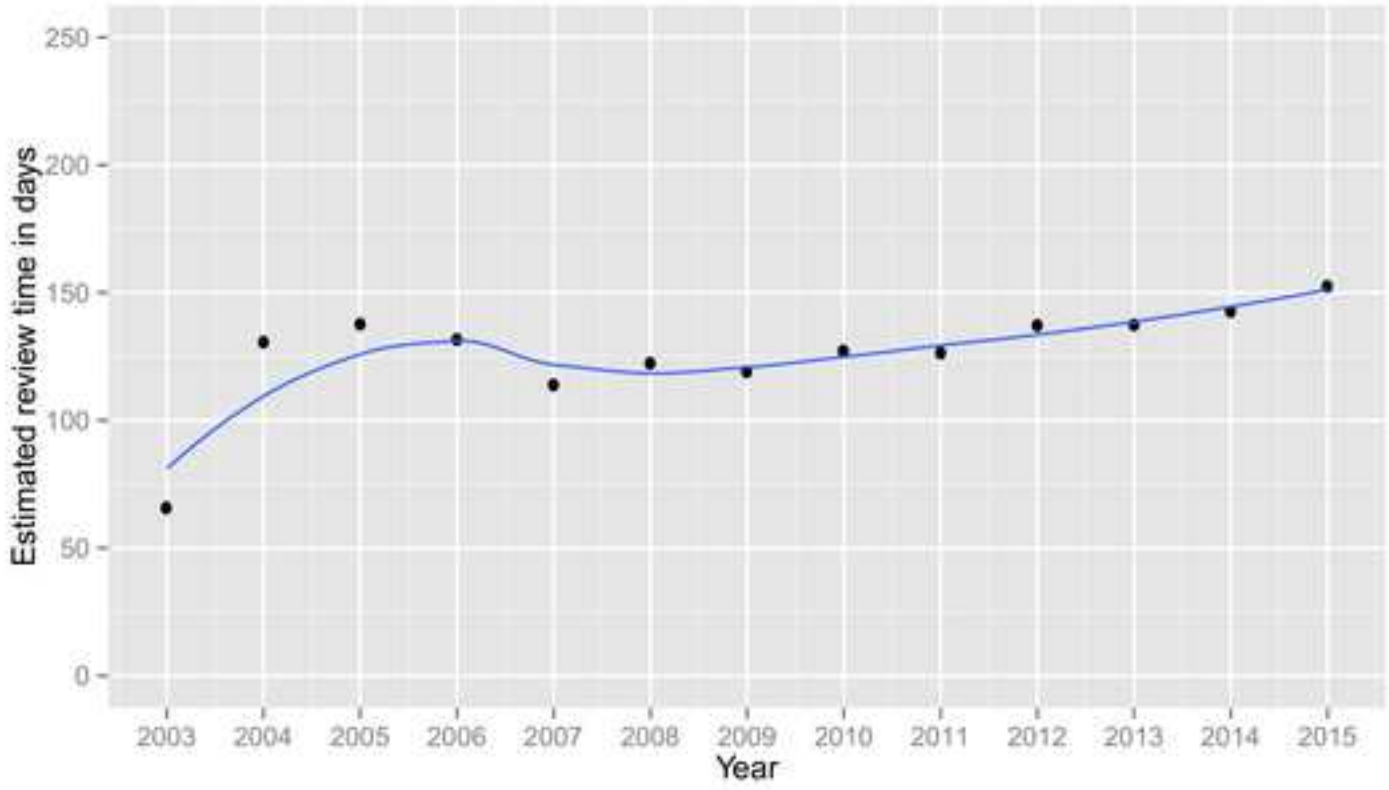
[8] R Core Team (2015) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Supporting Information

S1 Figure. Observed median review- and production time per journal.

S2 File. Poisson model estimates per journal for both review- and production time.

Fig 1
Click here to download Figure: Fig1.tif



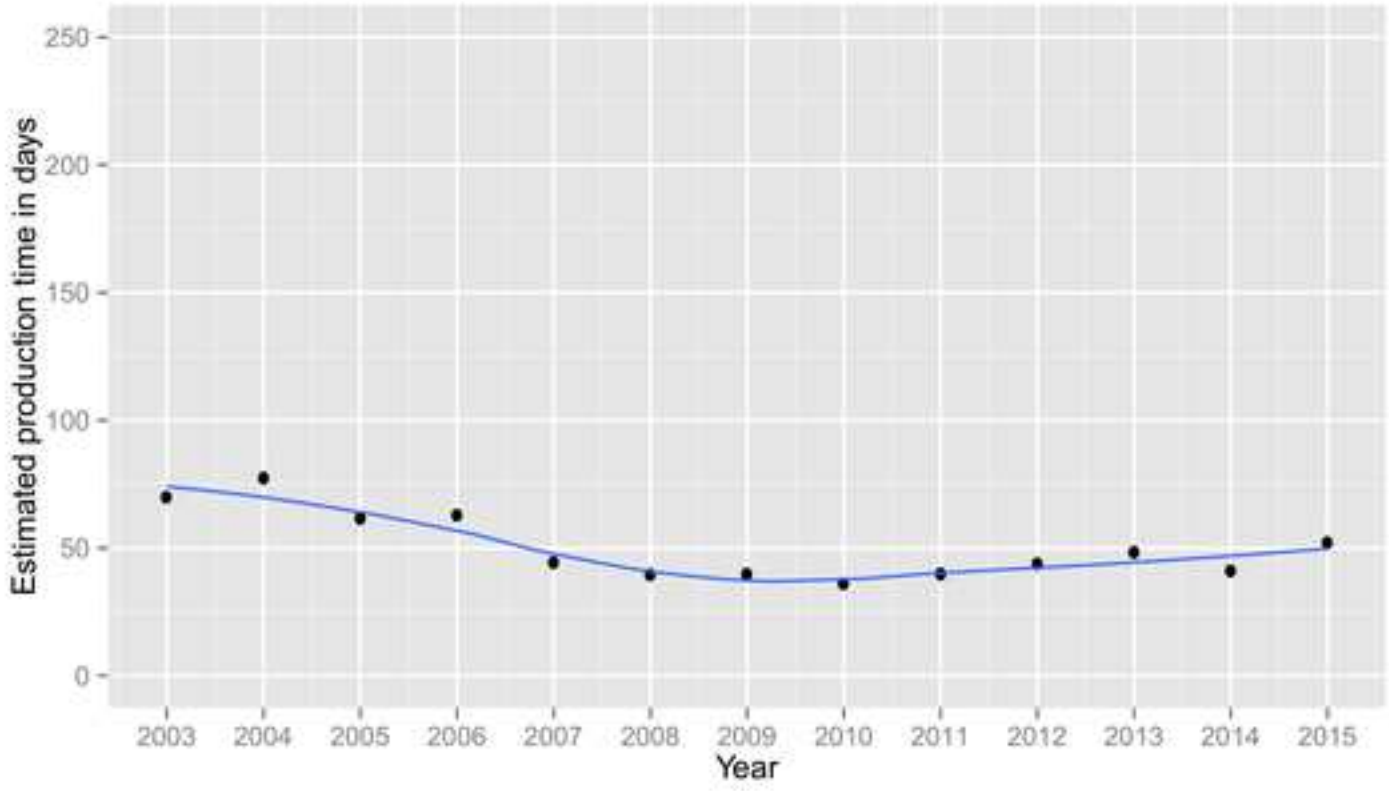
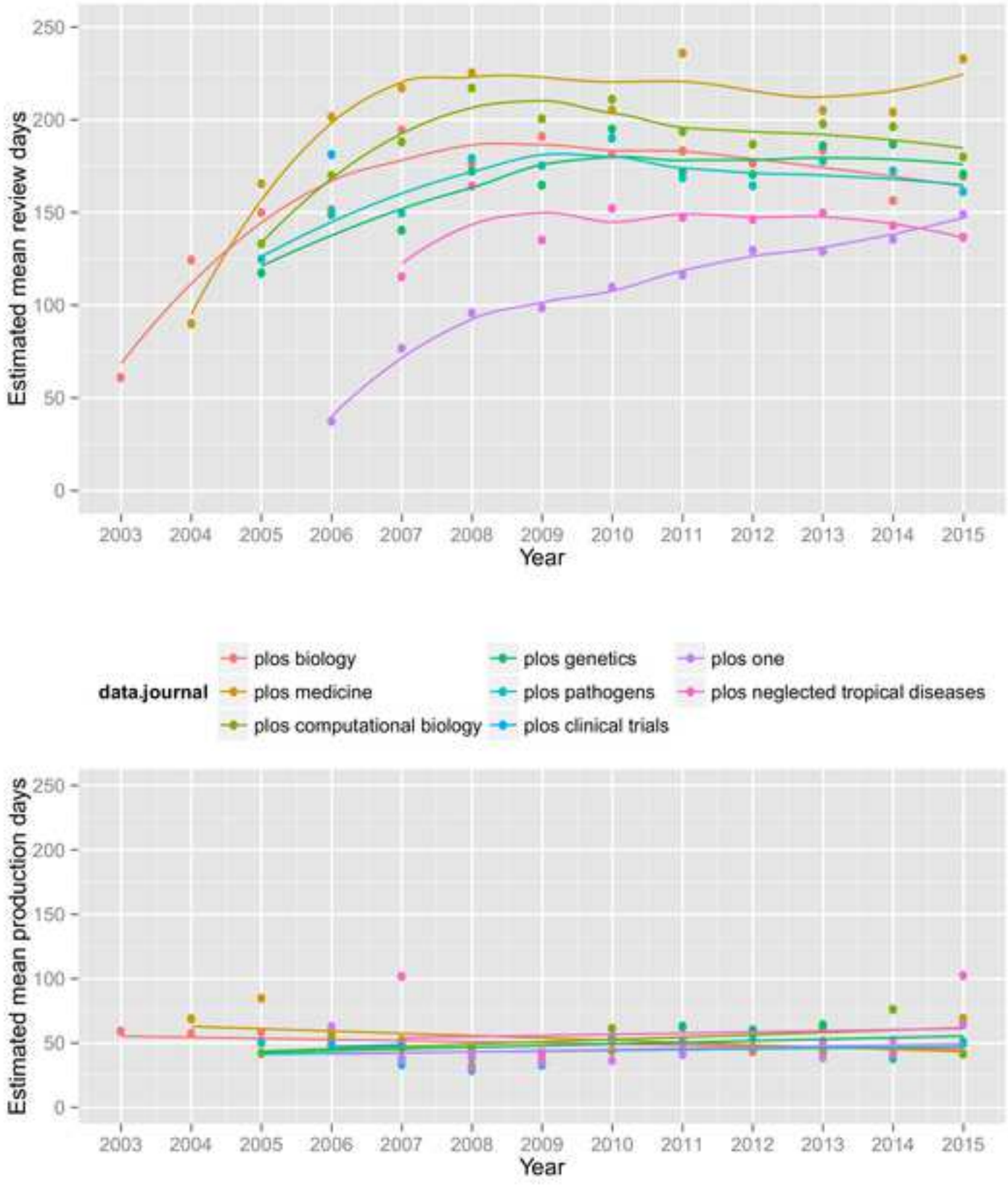


Fig 2
Click here to download Figure: Fig2.tif



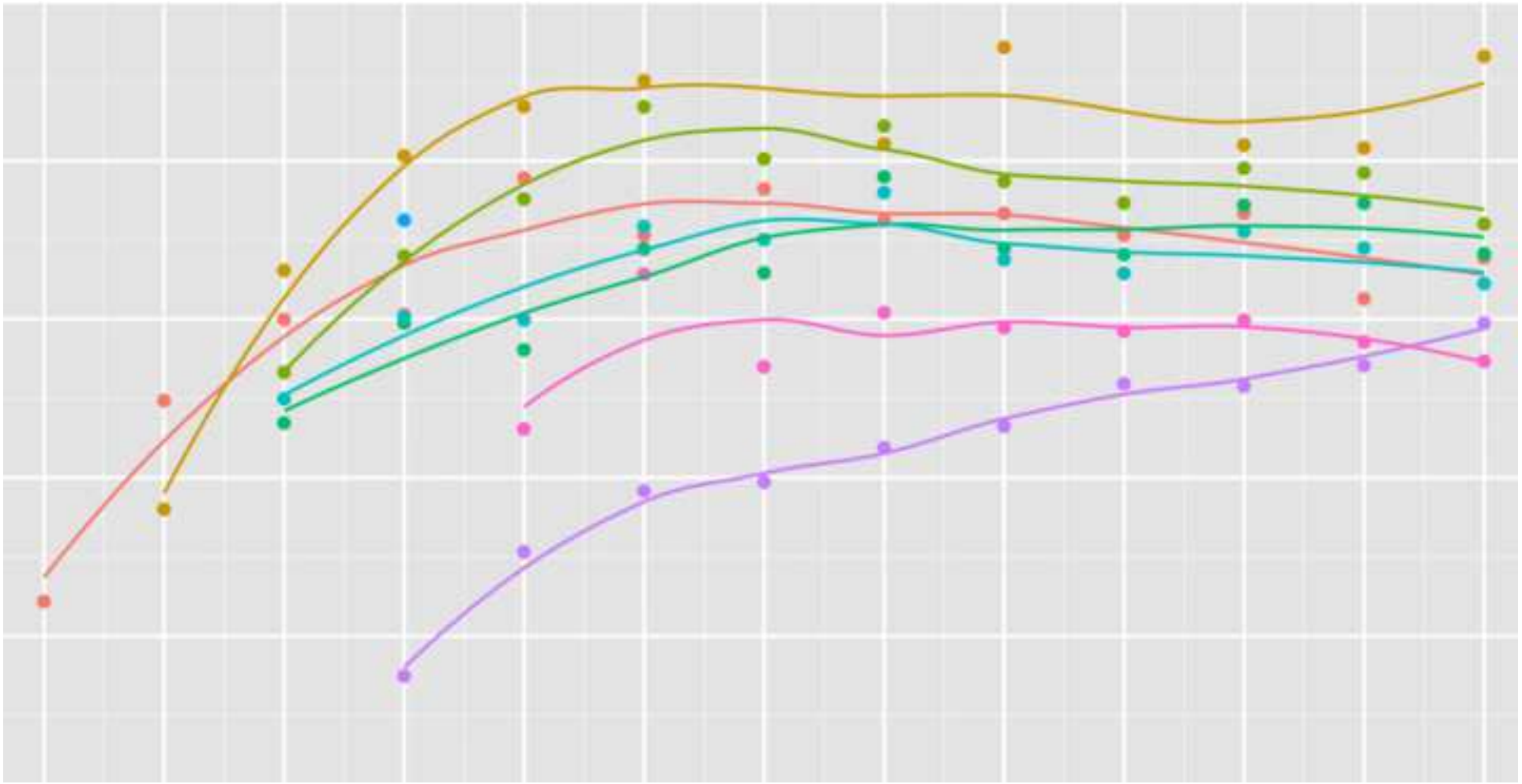
S1 Figure
Click here to download Supporting Information: S1 Observed median review and production times.pdf

S2 File

Click here to download Supporting Information: S2 Poisson model estimates per journal.xlsx

Striking Image

Click here to download Striking Image: striking image.tif



https://osf.io/53sn9/
http://osf.io/53sn9/