The inside track
Data analysis

Additional information for the News Feature
Nature **510**, 330–332 (2014); http://dx.doi.org/10.1038/510330a

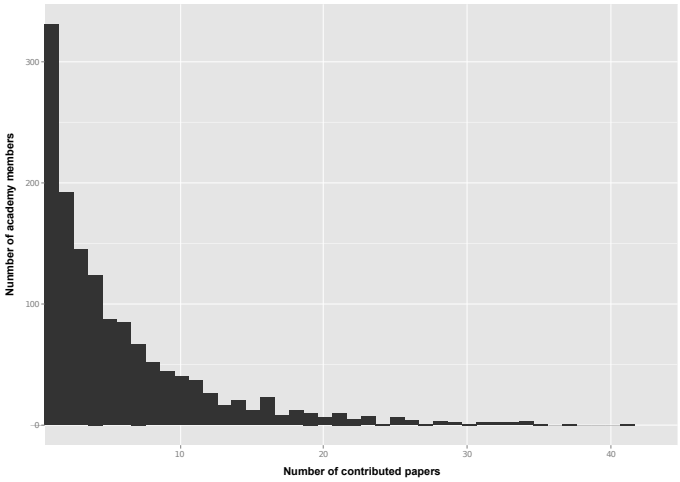

Figure 1

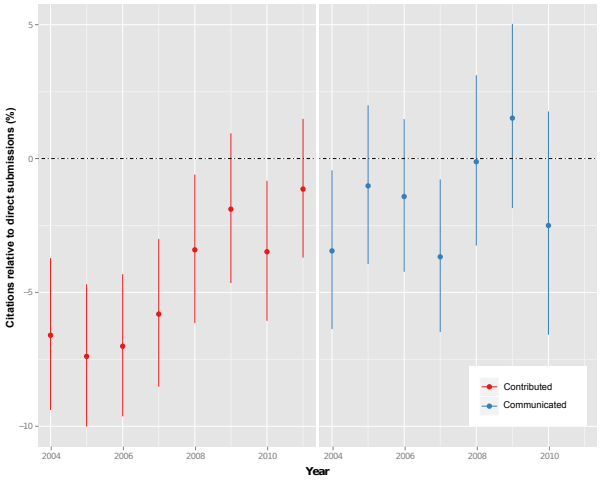

Figure 2

DATA

Python scripts were used to scrape metadata for abstracts at www.pnas.org for papers published in all issues of *Proceedings of the National Academy of Sciences* dated 2004–13. Recorded fields included: publication track (contributed, communicated and direct submissions); the individual who contributed, communicated or edited the paper; authors; dates defining the review period; title; online and print publication dates; whether the paper was open access; whether the paper had a prearranged editor; scientific discipline (defined by academy section); digital object identifier (DOI); and url.

The scraped data were cleaned with OpenRefine, using clustering algorithms and manual editing to standardize the name format for each academy member who contributed or communicated papers, and each editor, in the case of direct submissions. A small number of publications lacking academy-section labels, mostly not original research papers, were excluded from subsequent analyses. This yielded a total of 34,932 papers (22,408 direct submissions, 8,054 contributed and 4,470 communicated).

Python scripts were also used to scrape data on cumulative citations recorded by the CrossRef reference linking service, delivering citation counts for 34,409 of the papers in the cleaned data. For some analyses, citation data were also downloaded from the Thomson Reuters Web of Science. The two sources of citation data were strongly correlated (Pearson's $r = 0.955$, for a random sample of ~2,000 papers).

The cleaned data were managed and queried using an SQLite database. Statistical and graphical analyses were performed using R.

Data are available on request (peter@peteraldhous.com).

USE OF THE CONTRIBUTED TRACK

Over the decade, 1,389 academy members published at least one contributed paper. The full distribution shows that only a small group consistently published three or more contributed papers each year (Fig. 1).

The 910 living members elected to the academy between 2004 and 2013 were analysed further. Subsequent to their election, none contributed at a rate of more than three papers a year; 23 contributed at between two and three papers a year; 102 at between one and two papers a year; 343 at a rate of less than one paper a year; and 442 published no contributed papers.

CITATION ANALYSIS

Differences in citation rates between tracks were examined using linear regression models, with cumulative CrossRef citation counts as the dependent variable, transformed using the formula $\log_{10}(\text{citation count} + 1)$ to

TABLE 1

	Contributed papers			Communicated papers		
Period	**Effect size**	**95% confidence interval**	***p***	**Effect size**	**95% confidence interval**	***p***
2004–11	−4.43%	−5.39% to −3.47%	<0.001	−0.53%	−1.70% to 0.65%	0.37
2004	−6.60%	−9.39% to −3.72%	<0.001	−3.45%	−6.37% to −0.44%	<0.05
2005	−7.39%	−10.01% to −4.70%	<0.001	−1.02%	−3.94% to 1.99%	0.5
2006	−7.01%	−9.63% to −4.32%	<0.001	−1.42%	−4.23% to 1.47%	0.33
2007	−5.81%	−8.52% to −3.01%	<0.001	−3.67%	−6.48% to −0.78%	<0.05
2008	−3.41%	−6.14% to −0.60%	<0.05	−0.12%	−3.25% to 3.11%	0.94
2009	−1.89%	−4.65% to 0.94%	0.19	1.53%	−1.85% to 5.03%	0.38
2010	−3.48%	−6.06% to −0.83%	<0.05	−2.50%	−6.58% to 1.76%	0.25
2011	−1.14%	−3.70% to 1.48%	0.39	N/A	N/A	N/A

TABLE 2

	Contributed papers			Communicated papers		
Period	**Effect size**	**95% confidence interval**	***p***	**Effect size**	**95% confidence interval**	***p***
2004	−5.31%	−8.10% to −2.43%	<0.001	−3.03%	−5.92% to −0.04%	<0.05
2011	−2.01%	−4.45% to 0.50%	0.12	N/A	N/A	N/A

approximate a normal distribution. Exploratory analyses revealed that this transformation did not perform acceptably for recently published papers, so all subsequent citation analyses were restricted to papers published in the period 2004–11.

In addition to publication track, all regression models included the following explanatory variables: paper age (days from online publication to the date at which citations were counted); scientific discipline (academy section); and whether or not the paper was open access. For publication track, direct submissions were the reference group; for scientific discipline, the reference group was biochemistry, the academy section with the largest number of published papers.

In addition to analysing the period 2004–11, separate models were constructed for each year (Fig. 2, Table 1; effect sizes and confidence intervals are estimates from exponential back-transformation, relative to citations for direct submissions).

The narrowing of the gap in citations between contributed papers and direct submissions in recent years might have been an artefact of the shorter time available for newer papers to accumulate citations. So to investigate further, annual citation data were downloaded from the Web of Science for papers published in 2004 and 2011 and counts calculated for a citation window of matching duration (to 2006 for 2004 papers; to 2013 for 2011 papers). These counts were transformed as above, and regression models constructed as before, with paper age calculated relative to the end of the citation window. Results were broadly consistent with the analysis of cumulative CrossRef citations (Table 2).

TIME TO PUBLICATION

The speed of the contributed track is frequently mentioned by regular users as being central to its appeal. So for 32,197 papers for which the metadata included the date at which the paper was received for review (direct submissions and communicated) or sent for review (contributed), the subsequent lag in days to online publication was calculated, and survival curves were plotted (Fig. 3).

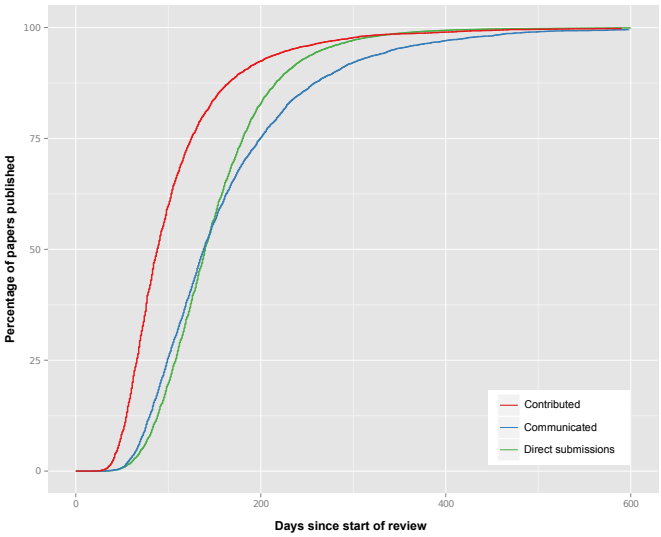

Figure 3

Note that lags for communicated papers include the time taken to select reviewers, and for direct submission include both this time and the time taken to select an editor for the paper. It is not possible from the data to determine the extent to which these periods explain the shorter publication lags for contributed papers.

ACKNOWLEDGEMENTS

Thanks to Phil Davis, David Rand and Anthony van Raan for advice on aspects of the citation analysis.