



# Identification and characterization of ‘self-promotion journals’: a survey of biomedical journals

Alexandre Scanff

## ► To cite this version:

Alexandre Scanff. Identification and characterization of ‘self-promotion journals’: a survey of biomedical journals. Life Sciences [q-bio]. 2020. dumas-03243394

**HAL Id: dumas-03243394**

**<https://dumas.ccsd.cnrs.fr/dumas-03243394>**

Submitted on 31 May 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



**THÈSE D'EXERCICE / UNIVERSITÉ DE RENNES 1**  
*sous le sceau de l'Université Bretagne Loire*

Thèse en vue du  
**DIPLÔME D'ÉTAT DE DOCTEUR EN MÉDECINE**

présentée par  
**Alexandre Scanff**

Né le 31 octobre 1989 (à Étampes)

---

**Identification et  
caractérisation  
"revues d'auto-  
promotion" : une  
étude des journaux  
biomédicaux**

**Thèse soutenue à Rennes  
le 20/10/2020**

devant le jury composé de :

**Jean-François Viel**

Professeur, CHU Rennes/ *Président du Jury*

**Ronan Garlantezec**

Professeur, CHU Rennes/ *examineur*

**Matthieu Revest**

Professeur, CHU Rennes/ *examineur*

**Clara Locher**

Docteur, CHU Rennes/ *directeur de thèse*

**Florian Naudet**

Professeur, CHU Rennes/ *co-directeur de thèse*

PROFESSEURS DES UNIVERSITES au 01/09/2020			
NOM	PRENOM	TITRE	SOUS-SECTION CNU
ANNE-GALIBERT	Marie-Dominique	PU-PH	Biochimie et biologie moléculaire
BARDOU-JACQUET	Edouard	PU-PH	Gastroentérologie ; hépatologie ; addictologie
BELAUD-ROTUREAU	Marc-Antoine	PU-PH	Histologie, embryologie et cytogénétique
BELLISSANT	Eric	PU-PH	Pharmacologie fondamentale ; pharmacologie clinique ; addictologie
BELOEIL	Hélène	PU-PH	Anesthésiologie-réanimation et médecine péri-opératoire
BENDAVID	Claude	PU-PH	Biochimie et biologie moléculaire
BENSALAH	Karim	PU-PH	Urologie
BEUCHEE	Alain	PU-PH	Pédiatrie
BONAN	Isabelle	PU-PH	Médecine physique et de réadaptation
BONNET	Fabrice	PU-PH	Endocrinologie, diabète et maladies métaboliques ; gynécologie médicale
BOUDJEMA	Karim	PU-PH	Chirurgie viscérale et digestive
BOUGET	Jacques	Professeur Emérite	Thérapeutique-médecine de la douleur ; addictologie
BOUGUEN	Guillaume	PU-PH	Gastroentérologie ; hépatologie ; addictologie
BRASSIER	Gilles	PU-PH	Neurochirurgie
BRETAGNE	Jean-François	Professeur Emérite	Gastroentérologie ; hépatologie ; addictologie
BRISOT	Pierre	Professeur Emérite	Gastroentérologie ; hépatologie ; addictologie
CARRE	François	Professeur Emérite	Physiologie
CATTOIR	Vincent	PU-PH	Bactériologie-virologie ; hygiène hospitalière
CHALES	Gérard	Professeur Emérite	Rhumatologie
COGNÉ	Michel	PU-PH	Immunologie
CORBINEAU	Hervé	PU-PH	Chirurgie thoracique et cardiovasculaire
CUGGIA	Marc	PU-PH	Biostatistiques, informatique médicale et technologies de communication
DAUBERT	Claude	Professeur Emérite	Cardiologie
DAVID	Véronique	PU-PH	Biochimie et biologie moléculaire
DAYAN	Jacques	Professeur Associé	Pédopsychiatrie ; addictologie
DE CREVOISIER	Renaud	PU-PH	Cancérologie ; radiothérapie
DECAUX	Olivier	PU-PH	Médecine interne ; gériatrie et biologie du vieillissement ; addictologie
DESRUES	Benoît	PU-PH	Pneumologie ; addictologie
DEUGNIER	Yves	Professeur Emérite	Gastroentérologie ; hépatologie ; addictologie
DONAL	Erwan	PU-PH	Cardiologie
DRAPIER	Dominique	PU-PH	Psychiatrie d'adultes ; addictologie
DUPUY	Alain	PU-PH	Dermato-vénérologie
ECOFFEY	Claude	PU-PH	Anesthésiologie-réanimation et médecine péri-opératoire
EDAN	Gilles	Professeur en surnombre	Neurologie
FERRE	Jean-Christophe	PU-PH	Radiologie et imagerie médicale
FEST	Thierry	PU-PH	Hématologie ; transfusion
FLECHER	Erwan	PU-PH	Chirurgie thoracique et cardiovasculaire
GANDEMER	Virginie	PU-PH	Pédiatrie
GANDON	Yves	PU-PH	Radiologie et imagerie médicale
GANGNEUX	Jean-Pierre	PU-PH	Parasitologie et mycologie
GARIN	Etienne	PU-PH	Biophysique et médecine nucléaire
GARLANTEZEC	Ronan	PU-PH	Epidémiologie, économie de la santé et prévention
GAUVRIT	Jean-Yves	PU-PH	Radiologie et imagerie médicale
GODEY	Benoît	PU-PH	Oto-rhino-laryngologie
GUGGENBUHL	Pascal	PU-PH	Rhumatologie
GUILLE	François	Professeur Emérite au 01/11/2020	Urologie
GUYADER	Dominique	PU-PH	Gastroentérologie ; hépatologie ; addictologie
HAEGELEN	Claire	PU-PH	Anatomie
HOUOT	Roch	PU-PH	Hématologie ; transfusion
JEGO	Patrick	PU-PH	Médecine interne ; gériatrie et biologie du vieillissement ; addictologie
JEGOUX	Franck	PU-PH	Oto-rhino-laryngologie
JOUNEAU	Stéphane	PU-PH	Pneumologie ; addictologie
KAYAL	Samer	PU-PH	Bactériologie-virologie ; hygiène hospitalière
LAMY DE LA CHAPELLE	Thierry	PU-PH	Hématologie ; transfusion
LAVIOLLE	Bruno	PU-PH	Pharmacologie fondamentale ; pharmacologie clinique ; addictologie
LAVOUE	Vincent	PU-PH	Gynécologie-obstétrique ; gynécologie médicale

NOM	PRENOM	TITRE	SOUS-SECTION CNU
LE BRETON	Hervé	PU-PH	Cardiologie
LE TULZO	Yves	PU-PH	Médecine intensive-réanimation
LECLERCQ	Christophe	PU-PH	Cardiologie
LEDERLIN	Mathieu	PU-PH	Radiologie et imagerie médicale
LEGUERRIER	Alain	Professeur Emérite	Chirurgie thoracique et cardiovasculaire
LE JEUNE	Florence	PU-PH	Biophysique et médecine nucléaire
LEVEQUE	Jean	PU-PH	Gynécologie-obstétrique ; gynécologie médicale
LIEVRE	Astrid	PU-PH	Gastroentérologie ; hépatologie ; addictologie
MABO	Philippe	PU-PH	Cardiologie
MAHE	Guillaume	PU-PH	Chirurgie vasculaire ; médecine vasculaire
MALLEDANT	Yannick	Professeur Emérite	Anesthésiologie-réanimation et médecine péri-opératoire
MATHIEU-SANQUER	Romain	PU-PH	Urologie
MENER	Eric	Professeur associé	Médecine générale
MICHELET	Christian	Professeur Emérite	Maladies infectieuses ; maladies tropicales
MOIRAND	Romain	PU-PH	Gastroentérologie ; hépatologie ; addictologie
MORANDI	Xavier	PU-PH	Anatomie
MOREL	Vincent	Professeur associé	Médecine palliative
MOSSER	Jean	PU-PH	Biochimie et biologie moléculaire
MOURIAUX	Frédéric	PU-PH	Ophthalmologie
MYHIE	Didier	Professeur associé	Médecine générale
NAUDET	Florian	PU-PH	Thérapeutique-médecine de la douleur ; addictologie
ODENT	Sylvie	PU-PH	Génétique
OGER	Emmanuel	PU-PH	Pharmacologie fondamentale ; pharmacologie clinique ; addictologie
PARIS	Christophe	PU-PH	Médecine et santé au travail
PERDRIGER	Aleth	PU-PH	Rhumatologie
PESCHANSKY	Nicolas	Professeur Associé	Médecine d'urgence
PLADYS	Patrick	PU-PH	Pédiatrie
RAVEL	Célia	PU-PH	Histologie, embryologie et cytogénétique
RENAUT	Pierric	Professeur associé	Médecine générale
REVEST	Matthieu	PU-PH	Maladies infectieuses ; maladies tropicales
RIFFAUD	Laurent	PU-PH	Neurochirurgie
RIOUX-LECLERCQ	Nathalie	PU-PH	Anatomie et cytologie pathologiques
ROBERT-GANGNEUX	Florence	PU-PH	Parasitologie et mycologie
ROPARS	Mickaël	PU-PH	Chirurgie orthopédique et traumatologique
SAINT-JALMES	Hervé	Professeur Emérite au 01/12/2020	Biophysique et médecine nucléaire
SAULEAU	Paul	PU-PH	Physiologie
SCHNELL	Frédéric	PU-PH	Physiologie
SEGUIN	Philippe	PU-PH	Anesthésiologie-réanimation et médecine péri-opératoire
SIPROUDHIS	Laurent	PU-PH	Gastroentérologie ; hépatologie ; addictologie
SOMME	Dominique	PU-PH	Médecine interne ; gériatrie et biologie du vieillissement ; addictologie
SOULAT	Louis	Professeur associé	Médecine d'urgence
SULPICE	Laurent	PU-PH	Chirurgie viscérale et digestive
TADIE	Jean Marc	PU-PH	Médecine intensive-réanimation
TARTE	Karin	PU-PH	Immunologie
TATTEVIN	Pierre	PU-PH	Maladies infectieuses ; maladies tropicales
THIBAUT	Ronan	PU-PH	Nutrition
THIBAUT	Vincent	PU-PH	Bactériologie-virologie ; hygiène hospitalière
THOMAZEAU	Hervé	Professeur Emérite au 01/11/2020	Chirurgie orthopédique et traumatologique
TORDJMAN	Sylvie	PU-PH	Pédopsychiatrie ; addictologie
VERHOYE	Jean-Philippe	PU-PH	Chirurgie thoracique et cardiovasculaire
VERIN	Marc	PU-PH	Neurologie
VIEL	Jean-François	PU-PH	Epidémiologie, économie de la santé et prévention
VIGNEAU	Cécile	PU-PH	Néphrologie
VIOLAS	Philippe	PU-PH	Chirurgie infantile
WATIER	Eric	PU-PH	Chirurgie plastique, reconstructrice et esthétique ; brûlologie
WODEY	Eric	PU-PH	Anesthésiologie-réanimation et médecine péri-opératoire

MAITRES DE CONFERENCES DES UNIVERSITES au 01/09/2020			
NOM	PRENOM	TITRE	SOUS-SECTION CNU
ALLORY	Emmanuel	MCF associé	Médecine générale
AME	Patricia	MCU-PH	Immunologie
AMIOT	Laurence	MCU-PH	Hématologie ; transfusion
ANSELM	Amédéo	MCU-PH	Chirurgie thoracique et cardiovasculaire
ARNAUD	Alexis	MCU-PH	Chirurgie infantile
BANATRE	Agnès	MCF associé	Médecine générale
BASTIAN	Benjamin	MCF associé	Médecine générale
BEGUE	Jean Marc	MCU-PH	Physiologie
BERTHEUIL	Nicolas	MCU-PH	Chirurgie plastique, reconstructrice et esthétique ; brûlologie
BROCHARD	Charlène	MCU-PH	Physiologie
CABILLIC	Florian	MCU-PH	Biologie cellulaire
CASTELLI	Joël	MCU-PH	Cancérologie ; radiothérapie
CAUBET	Alain	MCU-PH	Médecine et santé au travail
CHAPRON	Anthony	MCF	Médecine générale
CHHOR-QUENIART	Sidonie	MCF associé	Médecine générale
CORVOL	Aline	MCU-PH	Médecine interne ; gériatrie et biologie du vieillissement ; addictologie
DE TAYRAC	Marie	MCU-PH	Biochimie et biologie moléculaire
DEGEILH	Brigitte	MCU-PH	Parasitologie et mycologie
DROITCOURT	Catherine	MCU-PH	Dermato-vénéréologie
DUBOURG	Christèle	MCU-PH	Biochimie et biologie moléculaire
DUGAY	Frédéric	MCU-PH	Histologie, embryologie et cytogénétique
EDELIN	Julien	MCU-PH	Cancérologie ; radiothérapie
FIQUET	Laure	MCF associé	Médecine générale
GOUIN épouse THIBAUT	Isabelle	MCU-PH	Hématologie ; transfusion
GUILLET	Benoît	MCU-PH	Hématologie ; transfusion
JAILLARD	Sylvie	MCU-PH	Histologie, embryologie et cytogénétique
KALADJI	Adrien	MCU-PH	Chirurgie vasculaire ; médecine vasculaire
KAMMERER-JACQUET	Solène-Florence	MCU-PH	Anatomie et cytologie pathologiques
LAVENTU	Audrey	MCF	sciences physico-chimiques et ingénierie appliquée à la santé
LE GALL	François	MCU-PH	Anatomie et cytologie pathologiques
LEMAITRE	Florian	MCU-PH	Pharmacologie fondamentale ; pharmacologie clinique ; addictologie
MARTINS	Pédro Raphaël	MCU-PH	Cardiologie
MENARD	Cédric	MCU-PH	Immunologie
MICHEL	Laure	MCU-PH	Neurologie
MOREAU	Caroline	MCU-PH	Biochimie et biologie moléculaire
MOUSSOUNI	Fouzia	MCF	Informatique
NYANGO TIMOH	Krystel	MCU-PH	Anatomie
PANGAULT	Céline	MCU-PH	Hématologie ; transfusion
ROBERT	Gabriel	MCU-PH	Psychiatrie d'adultes ; addictologie
TURLIN	Bruno	MCU-PH	Anatomie et cytologie pathologiques
VERDIER épouse LORNE	Marie-Clémence	MCU-PH	Pharmacologie fondamentale ; pharmacologie clinique ; addictologie
ZIELINSKI	Agata	MCF	Philosophie

**Remerciements**

Je remercie le professeur Viel pour sa disponibilité et ses conseils tout au long de mon internat et pour avoir accepté de présider le jury pour cette thèse.

Je remercie le professeur Garlantezec pour sa paisibilité toujours bienvenue et sa considération envers tous les internes de Santé Publique.

Je remercie le professeur Revest pour avoir accepté d'apporter son concours à ce projet.

Je remercie le professeur Naudet pour son inébranlable dynamisme, ses idées foisonnantes et énergisantes au cours de toutes ces années, et pour avoir accepté de co-diriger cette thèse.

Je remercie le docteur Locher pour nous avoir lancés sur ce projet, pour sa rigueur tout au long de sa réalisation, et pour avoir accepté de co-diriger cette thèse.

## Table des matières

Liste des PU-PH et MCU-PH.....	2
Remerciements.....	5
Propos introductif.....	7
Abstract.....	8
Introduction.....	9
Methods.....	10
Data extraction.....	10
Indices calculation.....	10
Percentage of papers by the most prolific author and Gini index.....	10
Publication lag.....	11
Description of outlier journals.....	11
Analysis.....	11
Protocol changes.....	12
Results.....	13
Journals selection and description.....	13
Indices description.....	13
Percentage of papers by the most prolific author and Gini index.....	13
Field-specific variations.....	14
Publication lag.....	14
Description of outliers and identification of self-promotion journals.....	14
Discussion.....	16
References.....	19
Tables.....	20
Figures.....	25
Web Appendices.....	28

## Propos introductif

Ce travail a été mené dans la continuation d'une première étude réalisée par Clara Locher, Florian Naudet, David Moher et Ioana Cristea explorant les particularités du journal *New Microbes and New Infection* et la position singulière de Didier Raoult, supérieur hiérarchique de l'éditeur en chef, signataire de 32 % des articles publiés depuis la création du journal en 2013. Il a également attiré l'attention des auteurs après la publication d'articles acceptés en moins de 24h.

Ce premier travail avait évalué, sur les journaux dans le champ de l'infectiologie, si la captation d'une grande part des articles publiés par un unique auteur était une pratique fréquente. Il avait retrouvé une centralisation plus modeste des articles avec un 95<sup>ème</sup> percentile de la proportion des articles publiés par l'auteur le plus prolifique à 10 %, et principalement sur les journaux publiant peu d'articles.

Notre objectif ici était d'augmenter l'échelle d'analyse en explorant un plus grand nombre de journaux, et j'ai été intégré au projet afin d'apporter l'expertise technique en R en offrant une réduction algorithmique et un temps d'analyse réaliste. J'ai également pu apporter la proposition de la comparaison avec l'index de Gini, dans l'objectif de pouvoir mettre en évidence une inhomogénéité dans la diversité des auteurs et des voix de la recherche, en prenant en compte le nombre de publications de chaque auteur et plus seulement du plus prolifique.

En parallèle, Dorothy Bishop, professeur en neuro-psychologie développementale à l'université d'Oxford, a rédigé un article de blog explorant également les problématiques d'auteurs anormalement prolifiques en s'attachant plus fortement aux délais d'acceptation des articles pour identifier de potentiels passe-droits. La coïncidence étant trop belle, nous lui avons proposé et elle a accepté de participer avec nous à cette étude.

Le protocole en a été élaboré en concertation avec tous et déposé sur Open Science Framework en date du 21 juillet. La présente version de l'article a été réalisée et revue avec Clara Locher et Florian Naudet et reste une version intermédiaire, que nous allons proposer en parallèle aux trois autres collaborateurs pour relecture et approfondissement. La version qui sera présentée lors de la soutenance risque donc de présenter de légères variations par rapport à la version jointe.

J'espère que vous prendrez à la lire autant de plaisir que nous avons eu à l'écrire.



## **Identification and characterization of ‘self-promotion journals’: a survey of biomedical journals.**

### **Abstract**

**Context:** Converging analyses from different fields supports the idea that a high Percentage of Papers by the Most Prolific (PPMP) author is a simple measure that can be used as a red flag to identify “self-promotion” journals that may be suspected of biased editorial decision-making. The Gini index may also offer the ability to explore in-homogeneity in authorship volume related to more than one author.

**Methods:** For all biomedical journals referenced in the National Library of Medicine, with any attributed Broad Subject Terms, and at least 50 signed articles between 2015 and 2019, we automatically extracted all articles metadata and calculated the 2015-2019 PPMP and the 2015-2019 Gini index of distribution of articles among authors. When this information was available, we computed median of publication lag (delay between submission and acceptance) for articles with any of the most prolific author(s) and for article with none of the most prolific author(s). For outlier journals, defined as a PPMP or Gini index above the 95th percentile of their respective distribution, a random sample of 100 journals was described regarding impact factors, auto-citation and status among the editorial board of the most prolific author, among others.

**Results:** 5 468 journals having published 4 986 335 papers between 2015 and 2019 were analysed. PPMP 95<sup>th</sup> percentile was 10,6% (median was 2.9%). Gini index 95<sup>th</sup> percentile was 0,355 (median was 0.183). Correlation between indexes was 0,35 (95CI 0,33 to 0,37). Using a cut-off of 3 weeks for the median of publication lag, on 2 743 journals providing this information, 277 journals (10.2%) had a quicker median for articles with the most prolific author(s), versus 51 (1.9%) journals for articles without prolific author(s). Among the random sample of outlier journals, 98 provided information about their editorial board. Among those 98, the most prolific authors was part in the editorial board in 60 cases (61 %), among which 25 (26 % of the 98) were the editor in chief.

**Discussion:** Our results reveal relationships between authors and editors that question the necessary independence of scientific publishing. There may be defensible reasons for authors to participate in the editorial board of a journal where they largely publish but the high number of papers published by the PPMP authors and, in some cases, their shorter publication lag suggest a potential for corruption of the peer review process. Limits of this study included automated calculations with a risk of inaccuracy and a focus on journals indexed with NLM Broad Subject Terms that may have selected the more established journals.

**MeSH :** Bibliometrics, Editorial Policies, Publishing, Scientific integrity, Journalology

## Introduction

The concept of ‘self-promotion journals’ was proposed on the ground of the dubious editorial practices of New Microbes and New Infection (NMNI), an Elsevier journal, whose most prolific author, Didier Raoult co-authored 32% of all published papers. NMNI’s editor-in-chief is a subordinate of Raoult as well as six additional associate editors of the journal. Altogether, they authored 44% of the papers published up to June 25, 2020 in the journal. ‘Self-promotion journals’ were therefore proposed as a new type of illegitimate publishing entity may have certain key characteristics such as (i) a constantly high proportion of papers published by a group of authors, (ii) particularly in the presence of relationships between the editors and these authors, and (iii) publication of low-quality research. A preliminary approach applied to the field of infectious disease journals aimed to detect such ‘self-promotion journals’ using a simple measure: the proportion of contributions published in a journal by the most prolific author, i.e the one who published the most articles during a given period [1]. In journals publishing > 50 papers over 5 years, it was rare to see journals where a specific author published more than 10% of papers, and indeed NMNI was a clear outlier.

Coincidentally, Dorothy Bishop reported in a blogpost a similar analysis for the addiction subfield of psychology[2]. She found a bimodal distribution of ‘percent by most prolific’ measure, identical to the one was used for NMNI, with only 3 of 99 journals having a score over 8%. In two of these journals, the high score came from the same individual, who was on the editorial board of the journal, and who had published together with the editor-in-chief. Bishop also noted that the same method effectively identified a journal editor, Johnny Matson, who had been found previously to be publishing copiously in journals edited by himself or other editors without any peer review[2]. The absence of peer review was inferred by the remarkably rapid turnaround of papers, often with a week or less between the dates recorded between submission and acceptance. This confirmatory evidence of unethical editorial practice can only be obtained, however, in journals that report these dates on the published manuscript.

These converging analyses from different fields supports the idea that Percentage of Papers by the Most Prolific (PPMP) author is a simple measure that can be used as a red flag to identify journals that are suspected of biased editorial decision-making. Complementary, a more complex measure like the Gini index[3], largely used in econometry to describe resources distribution inequalities, offer the ability to explore inhomogeneity in authorship volume related to more than one author. Our objectives were (i) to apply the PPMP approach on a very large dataset of biomedical journals over a 5-year time frame, (ii) to contrast it with the Gini index, (iii) to describe more qualitatively outliers on these indices and (iv) to describe intervals between submission and acceptance dates of papers by the most prolific authors relative to papers with other authors.

## Methods

We developed and followed a research protocol that was prospectively registered the 21 July 2020 on the Open Science Framework (<https://osf.io/6e3uf/>). Code and summarized data are also available at the same URL.

### Data extraction

Eligibility criteria for the selection of journals were (i) to be a biomedical journal referenced in the National Library of Medicine (NLM), (ii) to have at least one ‘Broad Subject Term’ and, (iii) to have published more than 50 papers between January 2015 and December 2019. Broad Subject Terms are Medical Subject Headings (MeSH) terms used to describe a journal's overall scope and are defined by the NLM for journals in the MEDLINE database[4]. The 2015–2019 period was chosen as this 5-year window allows for smoothing of random variation yet describing recent practices. One author (AS) had search for journal change name during the 2015–2019 period and, in case of renaming, had pooled articles published under the different names.

To identify eligible journals, we used the Entrez programming utilities (E-utilities) that allow querying on the National Center for Biotechnology Information (NCBI) databases. The search query – presented in **Web-Appendix 1** – was used to identify all biomedical journals in the NLM catalogue having at least one of the ‘Broad Subject Term’ listed. Then, for each journal, article metadata were automatically collected with E-utilities. Due to technical restrictions, querying for article metadata was done from 2015 until the date of extraction and then restricted to January 2015 to December 2019.

Because of articles without an author name, the third selection criterion was slightly modified to focus on journal with at least 50 ‘signed articles’ – i.e. articles with at least one identified author – during the 2015–2019 period (see ‘protocol changes’). Publications on multiple journals (e.g. PRISMA statements, which was published in 6 different journals to promote its dissemination) did not receive special treatment, and no correction was applied as each article was only examined inside its journal.

### Indices calculation

#### *Percentage of papers by the most prolific author and Gini index*

For each journal, each author was identified by his or her full name (i.e. family name and complete forename) or, failing that, by his or her family name and initial(s) of forename. The number of articles signed by this author was count. The PPMP was defined as the number of articles of the most prolific author ( $n_{\max}$ ) divided by the total number of signed articles in the journal ( $N_{\text{tot}}$ ):  $\text{PPMP} = [n_{\max} / N_{\text{tot}}]$ . Complementary to the PPMP, the Gini index was used to explore inequality in authorship volume related to more than one author. The Gini index of number of appearances as an author was calculated among

articles with at least one author, with correction for total number of authors (see formula in **Web-Appendix 2**) [3]. The values of Gini index range between 0 (perfect equality) and 1 (total inequality). For the primary analysis, these two outcomes (PPMP and Gini) were computed for all papers, and as a sensitivity analysis, they were computed only for papers labelled as “journal articles” (using the NCBI publication type).

### *Publication lag*

For each article, the publication lag – defined as the time interval between submission and acceptance of an article – was computed when possible. Then, each journal was characterized by (i) median of publication lag for articles with any of the most prolific author(s) – in case of *ex aequo* on number of articles published – and (ii) median of publication lag for article with none of the most prolific author(s).

### **Description of outlier journals**

Outliers were defined as journals with a value of the PPMP and/or the Gini above the 95<sup>th</sup> percentile in the principal analysis. Among them, a sample of 100 journals was randomly selected. For random sampling, journals were first sorted by full name alphabetic order and randomly sampled using a random number generator with a seed arbitrarily set to 42 (R function `sample_n` in `dplyr` package,).

For these 100 outlier journals, one reviewer (AS or FN) scrutinized these journals and performed a manual extraction of characteristics related to journal impact factor (web of science: WoS), open access policies (WoS and Directory of Open Access Journals), open peer review policies (Publons), most prolific author h-index (WoS) and status on the editorial board of the journal (journal or publisher site). For the year 2019, the metrics ‘self-citation boost’ (i.e. number of self-citing articles over number of non-self-citing articles) and ‘skewness & nonarticle inflation’ (i.e. impact factor minus median of citation for an article, over median of citation for an article) was computed according to Ioannidis et Thombs[5].

### **Analysis**

A descriptive analysis was performed using median, range, and quartiles for continuous variable and counts and percentages for categorical variables. Description of the 100 outliers was realized overall and with respect to the membership of any of the most prolific author(s) to the editorial board. Correlations were computed using Pearson coefficient, with 95% confidence interval (CI).

To explore field-specific variations, distribution of the two indices inside each ‘Broad Subject term’ was graphically displayed. Yearly and global distribution of percentage of paper by each author and Lorenz curve – a graphical representation of the cumulative distribution of appearances as an author – were presented for each of the potential outliers identified above.

All analyses were realized using R version 3.6[6], and main packages RISmed 2.1 for queries of journal

characteristics[7], easyPubMed 2.13 for queries of article characteristics[8], DescTools 0.99 for Gini index calculation[9], and tidyverse 1.3 for miscellaneous[10].

### **Protocol changes**

Some practical challenges arose when performing our research as some articles unexpectedly lacked an author name, which preclude them to contribute to the numerator of PPMP or at all to the Gini index. Some journals also had more than 50 articles published from 2015 to 2019 but less than 50 had an author name. We therefore amended our definitions to make explicit that the PPMP denominator was defined as the number of articles with at least one identified author instead of all published article, and journals were included only if they have published 50 articles with author names instead of all published articles.

Description of the outlier journals with respect to the membership of any of the most prolific author(s) to the editorial board was added a posteriori for exploratory purpose.

## Results

### Journals selection and description

Using the search query, 11 665 journals labelled with at least one of 152 ‘Broad Subject Terms’ were retrieved. **Figure 1** details the reasons for non-inclusion of journals. Following the protocol change, 21 journals were excluded from the analysis because they have published more than 50 articles between 2015 and 2019 but less than 50 of them provided an author name. Finally, 5 468 journals were analysed. These journals published a total of 4 986 335 articles among which 4 582 473 were ‘journal article’ (see **Web-Appendix 3**). The main characteristics of the analysed journals are described in **Table 1**. Briefly, they published a median of 500 articles (IQR 262 to 964) for the 2015–2019 period, among which 463 (246 to 876) were labelled as ‘journal articles’. Two ‘mega-journals’ published more than 25 000 articles over the 5-year period, i.e. *Scientific Reports* with 95 890 signed articles, and *PloS One* with 107 342 signed articles. For 3 668 journals (67%), there was at least one article without author, with a median proportion of 0.9% of articles (IQR 0.4% to 2.1%) with no author name in these journals. The author with the highest number of articles in a given journal is a journalist (N = 767) and the author with the highest number of ‘journal articles’ is an academic (N = 471). Both write in a well-established journal and are parts of its editorial board.

### Indices description

#### *Percentage of papers by the most prolific author and Gini index*

For the 2015–2019 period, the PPMP median was 2.88% (IQR 1.71% to 4.91%) and the 95<sup>th</sup> percentile of PPMP value was 10.6% (**Figure 2**). The most prolific author(s) of each journal published a median of 14 articles (IQR 8 to 25). For 1 022 journals (19%), there was more than one author tied with the highest number of published articles. There were no sensible differences between indices values between years (**Web-Appendix 4**). There were no sensible differences between indices values between years, with 95<sup>th</sup> percentile of distribution ranging from 10.1% to 11.4% (**Web-Appendix 4**).

For the 2015–2019 period, the Gini index median was 0.183 (IQR 0.131 to 0.246), and the 95<sup>th</sup> percentile of Gini index was 0.355 (**Figure 2**). There were no sensible differences between indices values between years, with 95<sup>th</sup> percentile of distribution ranging from 0.212 to 0.224 (**Web-Appendix 4**). Correlation between PPMP and Gini indices was 0.35 (95CI 0.33 to 0.37).

Results of the sensitivity analyses based on ‘journal articles’ only were consistent (**Figure 2**). Correlations between indices computed for all articles or ‘journal articles’ only of respectively 0.94 (0.93 to 0.94) and 0.99 (0.99 to 0.99) for PPMP and Gini index.

For 540 journals (9.9%), at least a quarter of the authors only had initials of their forename(s) presented.

### *Field-specific variations*

Distribution of PPMP and Gini index among each NLM broad term are presented in **Web-Appendix 5**. Median PPMP per field ranged from 1.1% to 9.5%, with no field having a median superior to the 95th percentile threshold of the PPMP (i.e. 10.6%). Median Gini index per field ranged from 0.113 to 0.297, with no field having a median superior to the 95th percentile threshold of the Gini (i.e. 0.355).

### *Publication lag*

Because of missing data in all submission or acceptance dates, publication lag was not calculable for 2 743 journals (50.2%). These journals have fewer signed articles (369 (IQR 200 to 712) vs 637 (IQR 355 to 1 186)), no differences in Gini index but a higher PPMP (3.4% (2.0% to 5.9%) vs 2.4% (1.5% to 4.0%)) than journals which present submission and acceptance dates. For the remaining 2 725 (49.8%), the median of publication lags for all signed articles over five years is 85 days (IQR 53 to 123) for articles published by the most prolific author(s) versus 107 days (IQR 80 to 141) for articles published by none of the most prolific author(s).

**Figure 3** shows the scatter plot with marginal density curve of the median of publication lag for the most prolific author(s) versus non-prolific authors for each journal. For articles authored by the most prolific author(s), the distribution of the publication lag was skewed toward shorter lag. Using a cut-off of 3 weeks for the median of publication lag, 277 (10.2%) journals had a median inferior for articles with the most prolific author(s), 51 (1.9%) journals had a median inferior for articles without prolific author(s), and 38 (1.4%) journals had a median inferior for both articles (i.e. articles with or without the most prolific author(s)).

### **Description of outliers and identification of self-promotion journals**

Using the 95<sup>th</sup> percentile, we identified 480 outlier journals with respectively 206 with the PPMP only, 206 with the Gini only and 68 with the both indices. Yearly and global distribution of indices was presented for potential outliers identified above see (**Web-Appendix 6**).

Main characteristics of the 100 randomly selected outlier journals are presented in **Table 2**. Those 100 journals were written in English for 98 of them, among which 31 were also written in another language. The most common non-English languages were German (6 journals), Chinese (5 journals), Japanese (5 journals), French and Italian (4 journals each). These outliers were well-established journals, with a median year of start of activity in 1990 (IQR 1976 to 2001).

For 44 of these 100, they were not indexed in WoS, hampering any assessment of journal impact factor and other citation metrics. The 56 others presented a median journal impact factor of 2.9 (IQR 1.5 to 4.8) with a median of self-citation ratio of 0.11 (IQR 0.047 to 0.21), corresponding to a median of self-citing boost according to Ioannidis of 13% (IQR 5% to 26%). Ioannidis' skewness and non-article inflation

median was 86% (IQR 55% to 144%). Calculation of this metric was not possible for 11 journals (20%) that had a median of article citation of 0. Only 5 journals were indexed in the Directory of Open Access Journals as being full open access, and the median proportion of open access articles was 2.0% (0.47% to 8.0%). None of the outliers had an open peer review policy.

For 2 out of 100 journals, full composition of the editorial boards was not found and only the editor in chief were known and were not the most prolific author. In the 98 other journals, at least one of the most prolific authors was part in the editorial board in 60 cases (61%), among which 25 (26%) were the editor in chief. Journals where at least one of the most prolific authors was on the editorial board tend to have a higher impact factor with a median at 3.4 (IQR 2.0 to 5.3) than the other with a median at 1.4 (IQR 0.95 to 3.1).

In addition, for 14 cases, all of the most prolific author were absent from WoS (for 4) or identification was considered unreliable (for 10) due to homonyms, 4 with full forename and 6 with initials only. Among the 86 others journals, the median of the *h*-index of most prolific author was 29 (IQR 13 to 50).

Last, the careful look at 108 authors tied as most prolific in their respective journals, allowed to identify errors in authors identification for one journal, MMW-Fortschritte der Medizin (an outlier on the Gini index), where the most prolific “author” was named “Red”, which seems to be a diminutive for “Redaktion” and can thus encompass multiple physical individuals. Still, when ruling out ‘Red’ as a valid author name, the new most prolific author was a member of the editorial board, the PPMP went from 7.5% of 4 920 signed articles to 7.6% of 4 553 signed articles, and the Gini index went from 0.637 to 0.616. Out of 1 978 identified authors (except for ‘Red’), 1 435 (72.5%) did not published more than one article in this journal.



## Discussion

In this comprehensive survey of biomedical journals, we characterized several features of editor-author relationships. We found numerous journals where the most prolific author signed more than 10.6% of articles published in the journal. The threshold of 10.6% defined using the 95<sup>th</sup> percentile is very close to the threshold previously identified with the same method on the infectious disease journals[1]. It is also close to the 8% threshold identified on a subset of 99 addiction research journals citing at least 10 times articles by a prolific author in the field, as a natural separation in a bimodal distribution of the PPMP[2].

The correspondent 95<sup>th</sup> percentile of Gini index, computed over five years, was 0.355. We found no other studies using the Gini index to evaluate homogeneities in authorship except for collaborators or work field of an individual in a network analysis[11]. When it was computed per year, the 95<sup>th</sup> percentile of Gini index was smaller, at 0.20, suggesting that increasing the time-frame allow for more of occasional authors to be recruited while keeping the regular ones stables, increasing the apparent inhomogeneity. These two indexes explore complementary patterns of reduction of diversity in publications, PPMP tackling a unique author and Gini index being more sensible to a group of highly publishing authors but less able to find a unique problematic individual.

There was no evidence of field specific variation concerning the two indices across all included biomedical journals. On the subgroup of journals providing publication lag, the distribution for the most prolific author(s) was globally shorter and skewed toward shorter publication lag, suggesting that for certain journals, outliers on PPMP and/or Gini index or not, the peer review process was bypassed for the prolific authors. All these results were consistent when all articles were considered as well as when only ‘journal articles’ were considered, suggesting that editorials and correspondence are not the only drivers of the indicators we explored.

On a random sample of outlier journals identified using the PPMP and the Gini index, we found that the prolific authors can be an “established” scientists with a relatively high *h*-Index, and that 60% of these most prolific authors were editors in chief or members of the editorial board. Among these journals, half had no journal impact factor attributed by WoS but the other half had a median journal impact factor of 2.9 (IQR 1.5 to 4.8). These journals generally presented a large self-citation ratio.

These results reveal relationships between authors and editors that question the necessary independence of scientific publishing, especially for senior researchers and occasionally in journals with a “high” journal impact factor. Typically, academics are incentivised to publish on the ground of productivity based metrics. Any favoritism during the publishing process may advantage these prolific authors in the competition for funding, hiring, promotions and tenure. On the other hand, there may be defensible reasons for high impact author to participate in the editorial board of a journal were he or she largely publish[2] but the high number of papers published by the PPMP authors and, in some cases, the

publication lag inferior to 3 weeks for these prolific authors, suggest weaknesses and/or favoritism during the peer review process. This last result, albeit based on a sub-sample of journal, is crucial in providing evidence that editorial decisions were not only unusually quick but also selectively quick for the favored subset of prolific authors. It may have an impact on productivity-based metrics and suggest a risk of instrumentation, if not corruption, of the scientific enterprise, by using journals as a ‘publication laundry’ for ‘vanity publication’ [12] of authors closely related to the editorial board. Gaming productivity metrics have been widely described by using self-citation, honorary authorship and/or ghost writing. Manipulation of individuals metrics through a dedicated self-promotion journal level is a surely understudied way of gaming the system.

Importantly, such a descriptive and exploratory survey, based on a large available database provides information about the big picture of “self-promotion journals” but it may miss some more qualitative aspects, especially about quality of articles published in these journals. Quality of a scientific article is indeed a very difficult concept to measure and it cannot be easily summarized in quantitative metrics. In addition, we have restricted our analysis to journals indexed in the National Library of Medicine (NLM) under one or more existing broad terms. Still, some journals are registered without broad terms, which requires a manual pick-up by the NLM. This mean that our survey may have included the more established journals indexed in the NLM, that stood the test of time, with perhaps a better global and editorial conduct quality than non-indexed journals. Similarly, because we restricted our search to journals publishing a minimum of 50 papers between the 2015-2019, we may have missed smaller journals with less professional editorial practices.

More importantly, our automated calculations carries a risk of inaccuracy due to homonyms. Misidentification and/or merging of author names may bias the PPMP and the Gini index in both directions. The analysis of a random sample allowed for a disambiguation procedure by inspecting qualitatively the most prolific authors. Still, for 10 out of 96 journals where one of the most prolific author was referenced in WoS, we weren’t able to clearly rule out a merging of publications by homonyms, either because only the initials of the author forenames was known (6 cases), or because the forename was very common, like Julie or Ken (4 cases). The two proposed indicators, and their current calculation, may therefore not be used blindly but may rather serve as first step tools to screening for potentially problematic journals that may requires a careful description of their editorial practices.

While our results are exploratory hampering a widespread use of these indicators, we hope that further research may help to establish these easily computed indexes as a help for publishers, authors, and indeed scientific committees involved in promotion and tenure to screen for potentially biased journals, needing further investigations. Integrity based medicine are indeed needed to overcome the limitations of productivity-based medicine [13]. Any continuous efforts in precision and centralization of articles and journal meta-data regarding authorship, publication process, and standardization of article type

qualification may help to improve accuracy of the proposed indices. In addition, guidance for editors and publishers may be developed to delineate good practices from obvious misconduct. This will make it possible to propose more precise indicators, and more representative of what integrity in scientific publishing must be.

## References

- 1 Locher C, Moher D, Cristea I, *et al.* Publication by association: the Covid-19 pandemic reveals relationships between authors and editors. *MetaArXiv* 2020. doi:10.31222/osf.io/64u3s
- 2 Deevybee. BishopBlog: 'Percent by most prolific' author score: a red flag for possible editorial bias. BishopBlog. 2020.<http://deevybee.blogspot.com/2020/07/percent-by-most-prolific-author-score.html> (accessed 20 Jul 2020).
- 3 Cowell F. Measurement of inequality. In: *Handbook of Income Distribution*. Elsevier 2000. 87–166.
- 4 Broad Subject Terms for Indexed Journals. <https://wwwcf.nlm.nih.gov/serials/journals/index.cfm>
- 5 Ioannidis JPA, Thoms BD. A user's guide to inflated and manipulated impact factors. *Eur J Clin Invest* 2019;**49**:e13151. doi:10.1111/eci.13151
- 6 R Core Team. *R A Language and Environment for Statistical Computing*. Vienna, Austria: : R Foundation for Statistical Computing 2019. <https://www.R-project.org/>
- 7 Kovalchik S. *RISmed: Download Content from NCBI Databases*. 2017. <https://CRAN.R-project.org/package=RISmed>
- 8 Fantini D. *easyPubMed: Search and Retrieve Scientific Publication Records from PubMed*. 2019. <https://CRAN.R-project.org/package=easyPubMed>
- 9 Signorell A, Aho K, Alfons A, *et al.* *DescTools: Tools for Descriptive Statistics*. 2020. <https://CRAN.R-project.org/package=DescTools>
- 10 Wickham H, RStudio. *tidyverse: Easily Install and Load the 'Tidyverse'*. 2017. <https://CRAN.R-project.org/package=tidyverse>
- 11 Chien T-W, Chang Y, Wang H-Y. Understanding the productive author who published papers in medicine using National Health Insurance Database: A systematic review and meta-analysis. *Medicine (Baltimore)* 2018;**97**:e9967. doi:10.1097/MD.0000000000009967
- 12 Grandjean P. Science for Precautionary Decision-Making. In: *Late Lessons from Early Warnings: Science, Precaution, Innovation*. Copenhagen: European Environment Agency 2013. 517–35.<https://dash.harvard.edu/handle/1/32306628>
- 13 Moher D, Naudet F, Cristea IA, *et al.* Assessing scientists for hiring, promotion, and tenure. *PLOS Biol* 2018;**16**:e2004089. doi:10.1371/journal.pbio.2004089

**Tables**

	<b>All articles</b>	<b>Journal Article</b>
Number of journals with $\geq 50$ signed articles	5468	5455
Number of articles		
Median [IQR]	500 [262 – 964]	463 [246 – 876]
Range	50 – 108 990	50 – 105 031
Number of articles with an author		
Median [IQR]	494 [257 – 952]	458 [243 – 868]
Range	50 – 107 342	50 – 104 713
PPMP (%)		
Median [IQR]	2.9 [1.7 – 4.9]	2.7 [1.6 – 4.5]
Range	0.1 – 39.9	0.1 – 42.7
95th percentile	10.6	9.6
Number of articles by MPA		
Median [IQR]	14 [8 – 25]	12 [8 – 21]
Range	1 – 767	1 – 471
More than one author tied as MPA	1 022 (19%)	1 156 (21%)
Gini index		
Median [IQR]	0.183 [0.131 – 0.246]	0.172 [0.123 – 0.232]
Range	0.000 – 0.741	0.000 – 0.715
95th percentile	0.355	0.343
Median of publication lag ratio (MPA / no MPA)		
Median [IQR]	0.83 [0.61 – 1.02]	0.86 [0.65 – 1.03]
Range	0 – 26.75	0 – 7.66
Not calculable	2 743 (50%)	2 768 (51%)

**Table 1:** Main characteristics of all journals in the United States National Library of Medicine catalog having at least one Broad Subject term and having published at least 50 signed articles between 2015 and 2019.

	<b>Total (N=100)</b>	<b>None of the most prolific authors in the editorial board (N=38)</b>	<b>At least one of the most prolific authors in the editorial board (N=60) *</b>
Starting activity year			
Median [IQR]	1990 [1976 – 2001]	1984 [1973 – 1998]	1994 [1982 – 2002]
Range	1879 – 2013	1879 – 2009	1905 – 2013
Number of articles			
Median [IQR]	425 [202 – 1 015]	334 [132 – 906]	440 [243 – 1 082]
Range	50 – 12 584	52 – 12 584	50 – 8 075
Number of articles with an author			
Median [IQR]	421 [197 – 1 013]	331 [132 – 906]	436 [240 – 1 065]
Range	50 – 12 555	50 – 12 555	50 – 8 075
Country			
United States	33 (33%)	14 (37%)	19 (32%)
England	16 (16%)	3 (7.9%)	13 (22%)
Japan	10 (10%)	6 (16%)	3 (5%)
Germany	7 (7%)	2 (5%)	5 (8%)
Italy	6 (6%)	2 (5%)	4 (7%)
Other	28 (28%)	11 (29%)	16 (27%)
Only initials of forename (for the most prolific author)			
No	87 (87%)	34 (89%)	52 (87%)
Yes	13 (13%)	4 (11%)	8 (13%)
Only initials of forename ( % of all authors)			
Median [IQR]	1.3 [0.4 – 7.0]	0.97 [0.17 – 3.8]	1.6 [0.53 – 9.1]
Range	0.0 – 95.7	0.0 – 94.1	0.0 – 95.7
PPMP (%)			
Median [IQR]	10.9 [5.1 – 14.7]	11.1 [5.2 – 14.7]	10.9 [5.2 – 14.7]
Range	0.7 – 34.8	0.7 – 34.8	1.0 – 23.8
Number of authors having published more than 10.6 % of all articles			
0	45 (45%)	16 (42%)	28 (47%)
1	38 (38%)	10 (26%)	28 (47%)
2 and more	17 (17%)	12 (32%)	4 (7%)
Number of articles by MPA			
Median [IQR]	39 [23 – 61]	30 [20 – 59]	45 [25 – 64]
Range	6 – 367	8 – 161	6 – 367
PPMP (on ‘journal articles’ only) (%)			
Median [IQR]	9.4 [4.8 – 14.4]	11.3 [5.0 – 14.8]	8.8 [4.8 – 12.5]
Range	0.7 – 34.7	0.7 – 34.7	1.0 – 22.7
Less than 50 signed ‘journal articles’	2 (2%)	1 (3%)	1 (2%)

	<b>Total (N=100)</b>	<b>None of the most prolific authors in the editorial board (N=38)</b>	<b>At least one of the most prolific authors in the editorial board (N=60) *</b>
Number of authors having published more than 9.6 % of all ‘journal articles’			
0	51 (51%)	16 (42%)	34 (57%)
1	23 (23%)	6 (16%)	17 (28%)
2 and more	24 (24%)	15 (39%)	8 (13%)
Less than 50 signed ‘journal articles’	2 (2%)	1 (3%)	1 (2%)
Number of articles by MPA (on ‘journal articles’ only)			
Median [IQR]	31 [19 – 56]	28 [20 – 59]	34 [21 – 56]
Range	8 – 321	8 – 161	8 – 321
Less than 50 signed ‘journal articles’	2 (2%)	1 (3%)	1 (2%)
Gini index			
Median [IQR]	0.363 [0.266 – 0.408]	0.367 [0.292 – 0.410]	0.359 [0.244 – 0.408]
Range	0.074 – 0.717	0.130 – 0.717	0.074 – 0.637
Gini index (on ‘journal articles’ only)			
Median [IQR]	0.352 [0.244 – 0.391]	0.358 [0.279 – 0.408]	0.329 [0.232 – 0.387]
Range	0.070 – 0.715	0.127 – 0.715	0.070 – 0.575
Less than 50 signed ‘journal articles’	2 (2%)	1 (3%)	1 (2%)
Position on the editorial board of the MPA			
No	38 (38%)	38 (100%)	0 (0%)
Yes	35 (35%)	0 (0%)	35 (58%)
Editor-in-Chief	25 (25%)	0 (0%)	25 (42%)
(Missing)	2 (2%)	0 (0%)	0 (0%)
Highest h-index of the MPA			
Median [IQR]	29 [13 – 50]	26 [11 – 36]	30 [15 – 53]
Range	1 – 161	1 – 161	2 – 122
Unreliable MPA	10 (10%)	6 (16%)	3 (5%)
MPA not found	4 (4%)	2 (5%)	2 (3%)
Journal impact factor (2014–2019)			
Median [IQR]	2.89 [1.54 – 4.82]	1.39 [0.95 – 3.07]	3.38 [2.04 – 5.34]
Range	0.63 – 60.80	0.63 – 30.33	0.79 – 60.80
(Missing)	46 (46%)	20 (53%)	24 (40%)
Journal impact factor (2019)			
Median [IQR]	2.42 [1.29 – 4.95]	1.38 [1.07 – 3.02]	3.12 [1.73 – 5.53]
Range	0.64 – 64.8	0.64 – 27.6	0.77 – 64.8
(Missing)	44 (44%)	20 (53%)	22 (37%)

	<b>Total (N=100)</b>	<b>None of the most prolific authors in the editorial board (N=38)</b>	<b>At least one of the most prolific authors in the editorial board (N=60) *</b>
<b>Journal impact factor (2019) without self-citations</b>			
Median [IQR]	2.15 [1.26 – 4.28]	1.28 [0.96 – 2.92]	2.44 [1.61 – 4.85]
Range	0.50 – 63.91	0.58 – 26.77	0.50 – 63.91
(Missing)	44 (44%)	20 (53%)	22 (37%)
<b>Self-citation ratio</b>			
Median [IQR]	0.114 [0.047 – 0.207]	0.063 [0.025 – 0.184]	0.128 [0.050 – 0.206]
Range	0.000 – 0.396	0.000 – 0.396	0.000 – 0.354
(Missing)	44 (44%)	20 (53%)	22 (37%)
<b>Self-citing boost (%)</b>			
Median [IQR]	12.9 [4.9 – 26.1]	6.8 [2.6 – 22.8]	14.6 [5.3 – 26.0]
Range	0.0 – 65.6	0.0 – 65.6	0.0 – 54.7
(Missing)	44 (44%)	20 (53%)	22 (37%)
<b>Article citation median</b>			
Median [IQR]	1 [1 – 3]	1 [0 – 2]	2 [1 – 3]
Range	0 – 42	0 – 20	0 – 42
(Missing)	45 (45%)	20 (53%)	23 (38%)
<b>Skewness and nonarticle inflation (%)</b>			
Median [IQR]	69.3 [49.6 – 97.4]	53.9 [35.2 – 88.2]	80.4 [54.8 – 99.4]
Range	-2.6 – 266.2	-2.60 – 214.1	12.9 – 266.2
(Missing)	56 (56%)	27 (71%)	27 (45%)
<b>Percentage of citable in gold open acces (2017–2019)</b>			
Median [IQR]	2.0 [0.5 – 8.0]	0.67 [0.0 – 4.9]	2.1 [1.0 – 11.4]
Range	0.0 – 100.0	0.0 – 100.0	0.0 – 100.0
(Missing)	44 (44%)	20 (53%)	22 (37%)
<b>Journal indexed in DOAJ</b>			
No	95 (95%)	37 (97%)	56 (93%)
Yes	5 (5%)	1 (3%)	4 (7%)
<b>Median of publication lag (All articles) (days)</b>			
Median [IQR]	78 [39 – 99]	94 [52 – 117]	66 [37 – 85]
Range	4 – 186	33 – 186	4 – 115
Not calculable	69 (69%)	28 (74%)	39 (65%)
<b>Median of publication lag (With any MPA) (days)</b>			
Median [IQR]	59 [38 – 92]	76 [39 – 90]	46 [38 – 91]
Range	6 – 212	6 – 212	15 – 120
Not calculable	73 (73%)	29 (76%)	42 (70%)

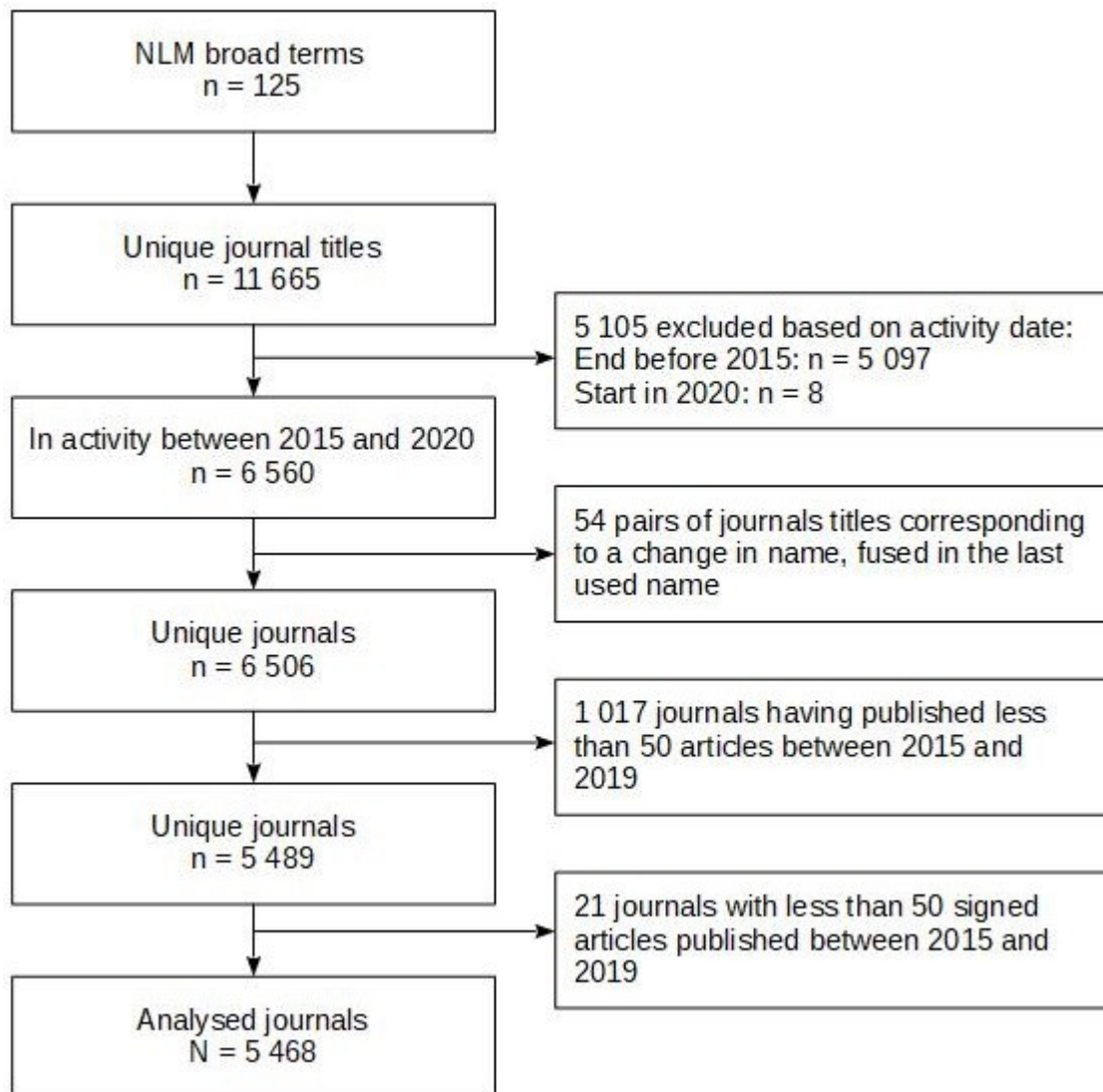


	<b>Total (N=100)</b>	<b>None of the most prolific authors in the editorial board (N=38)</b>	<b>At least one of the most prolific authors in the editorial board (N=60) *</b>
Median of publication lag (Without any MPA) (days)			
Median [IQR]	80 [41 – 99]	95 [55 – 118]	63 [36 – 83]
Range	4 – 184	33 – 184	4 – 115
Not calculable	69 (69%)	28 (74%)	39 (65%)
Median of publication lag ratio (MPA / no MPA) (days)			
Median [IQR]	0.90 [0.67 – 1.16]	0.77 [0.61 – 1.08]	0.90 [0.73 – 1.20]
Range	0.13 – 3.18	0.13 – 1.18	0.27 – 3.18
Not calculable	73 (73%)	29 (76%)	42 (70%)

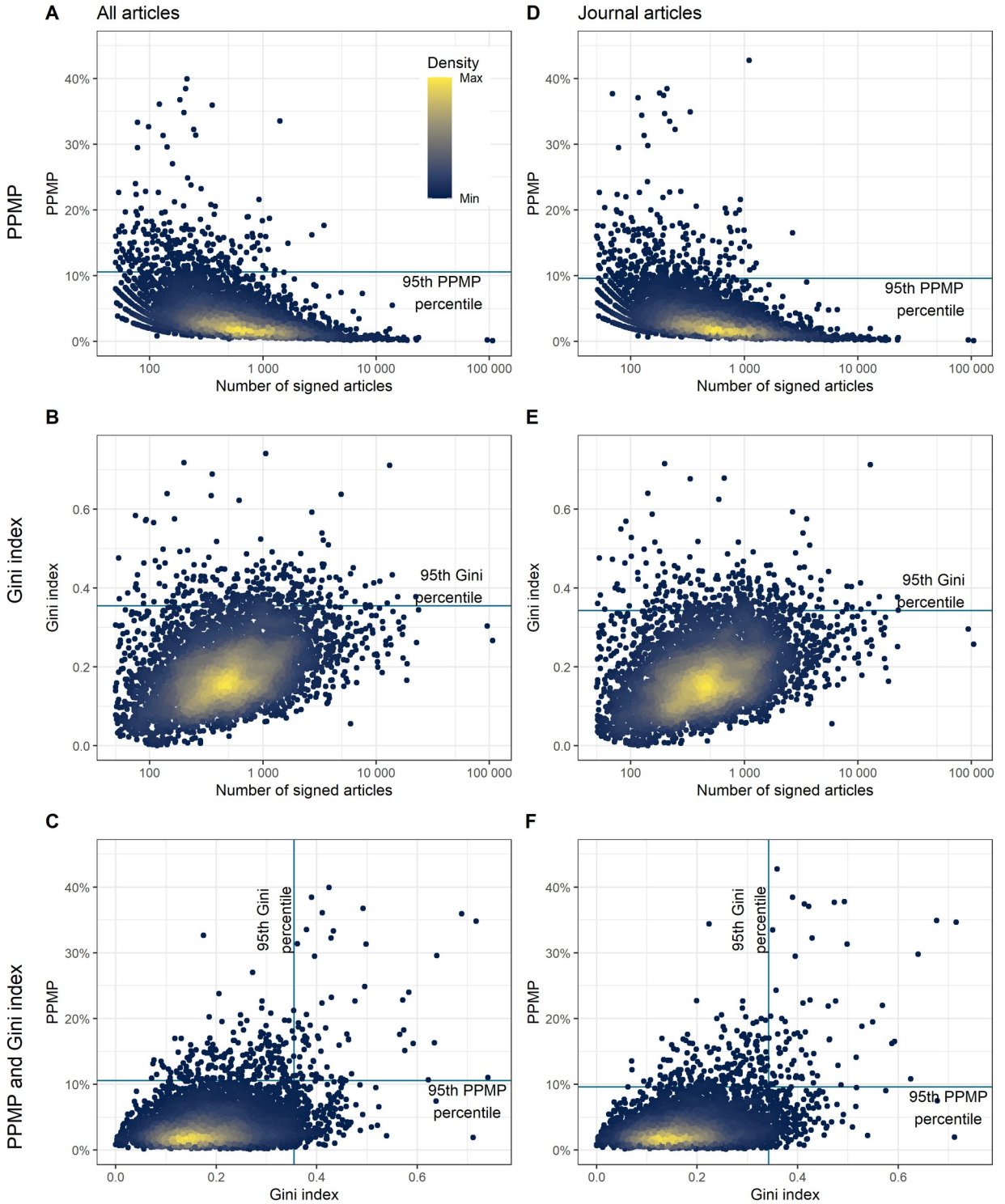
**Table 2:**

Main characteristics of 100 randomly selected journals with a PPMP or a Gini index higher than the 95<sup>th</sup> percentile, among the journals in the United States National Library of Medicine catalog having at least one Broad Subject term and having published at least 50 signed articles between 2015 and 2019. \* For two journals, position on the editorial board of the most prolific author could not be established.

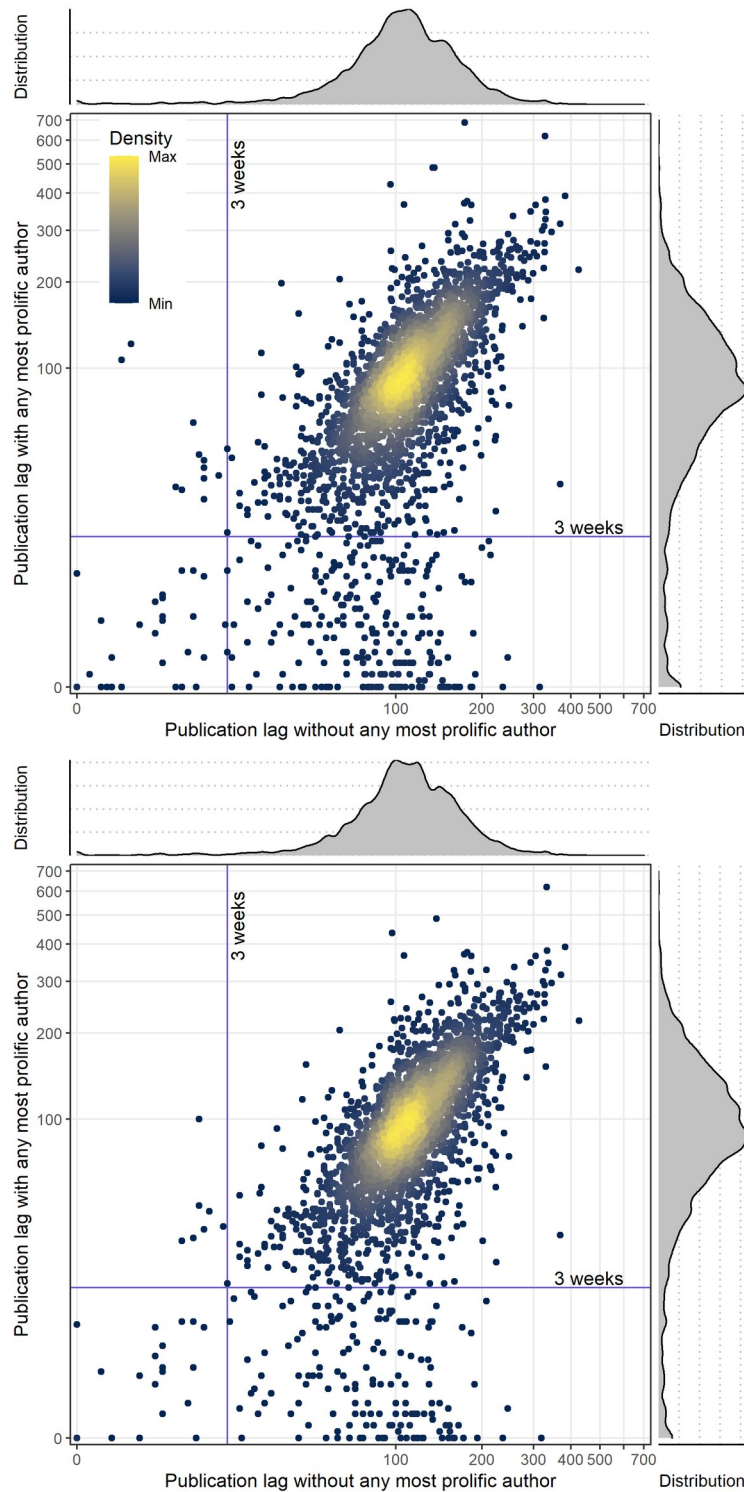
## Figures



**Figure 1:** Selection flowchart of journals labelled with at least one ‘Broad Subject Term’ in the National Library of Medicine (NLM).



**Figure 2:** Distribution of Percentage of Papers by the Most Prolific (PPMP) author(s) (A and D) and Gini index (B and E) compared to journal size, and comparison between PPMP and Gini index (C and F), among all articles (A to C) and among articles labelled as ‘Journal Articles’ only (D to E), published by all journals in the United States National Library of Medicine catalog having at least one Broad Subject term and having published at least 50 signed articles between 2015 and 2019.



**Figure 3:** Distribution of publication lag (in days) for articles with most prolific author compared to the ones without any, among all articles (top) and among articles labelled as "Journal Articles" only (bottom), published by all journals in the United States National Library of Medicine catalog having at least one Broad Subject term and having published at least 50 signed articles between 2015 and 2019.. With marginal density plot of distributions.

## Web Appendices

### Web-Appendix 1

United States National Library of Medicine catalog journal identification query:

("acquired immunodeficiency syndrome"[st] OR "aerospace medicine"[st] OR "allergy and immunology"[st] OR "anatomy"[st] OR "anesthesiology"[st] OR "anthropology"[st] OR "anti infective agents"[st] OR "antineoplastic agents"[st] OR "audiology"[st] OR "bacteriology"[st] OR "behavioral sciences"[st] OR "biochemistry"[st] OR "biology"[st] OR "biomedical engineering"[st] OR "biophysics"[st] OR "biotechnology"[st] OR "botany"[st] OR "brain"[st] OR "cardiology"[st] OR "cell biology"[st] OR "chemistry"[st] OR "chemistry techniques, analytical"[st] OR "chemistry, clinical"[st] OR "chiropractic"[st] OR "clinical laboratory techniques"[st] OR "communicable diseases"[st] OR "complementary therapies"[st] OR "computational biology"[st] OR "critical care"[st] OR "dentistry"[st] OR "dermatology"[st] OR "diagnostic imaging"[st] OR "disaster medicine"[st] OR "drug therapy"[st] OR "education"[st] OR "embryology"[st] OR "emergency medicine"[st] OR "endocrinology"[st] OR "environmental health"[st] OR "epidemiology"[st] OR "ethics"[st] OR "family planning services"[st] OR "forensic sciences"[st] OR "gastroenterology"[st] OR "general surgery"[st] OR "genetics"[st] OR "genetics, medical"[st] OR "geriatrics"[st] OR "gynecology"[st] OR "health services"[st] OR "health services research"[st] OR "hematology"[st] OR "histocytochemistry"[st] OR "histology"[st] OR "history of medicine"[st] OR "hospitals"[st] OR "internal medicine"[st] OR "jurisprudence"[st] OR "laboratory animal science"[st] OR "library science"[st] OR "medical informatics"[st] OR "medicine"[st] OR "metabolism"[st] OR "microbiology"[st] OR "military medicine"[st] OR "molecular biology"[st] OR "nanotechnology"[st] OR "neoplasms"[st] OR "nephrology"[st] OR "neurology"[st] OR "neurosurgery"[st] OR "nuclear medicine"[st] OR "nursing"[st] OR "nutritional sciences"[st] OR "obstetrics"[st] OR "occupational medicine"[st] OR "ophthalmology"[st] OR "orthodontics"[st] OR "orthopedics"[st] OR "osteopathic medicine"[st] OR "otolaryngology"[st] OR "palliative care"[st] OR "parasitology"[st] OR "pathology"[st] OR "pediatrics"[st] OR "perinatology"[st] OR "pharmacology"[st] OR "pharmacy"[st] OR "photography"[st] OR "physical and rehabilitation medicine"[st] OR "physics"[st] OR "physiology"[st] OR "podiatry"[st] OR "primary health care"[st] OR "psychiatry"[st] OR "psychology"[st] OR "psychopharmacology"[st] OR "psychophysiology"[st] OR "public health"[st] OR "pulmonary medicine"[st] OR "radiology"[st] OR "radiotherapy"[st] OR "reproductive medicine"[st] OR "rheumatology"[st] OR "science"[st] OR "sexually transmitted diseases"[st] OR "social sciences"[st] OR "speech language pathology"[st] OR "sports medicine"[st] OR "statistics as topic"[st] OR "substance related disorders"[st] OR "technology"[st] OR "teratology"[st] OR "therapeutics"[st] OR "toxicology"[st] OR "transplantation"[st] OR "traumatology"[st] OR "tropical medicine"[st] OR "urology"[st] OR "vascular diseases"[st] OR "veterinary medicine"[st] OR "virology"[st] OR "vital statistics"[st] OR "women's health"[st] OR "zoology"[st])

### Web-Appendix 2

Gini index calculation formula:

$$\left[ \frac{2 \sum y_i}{n \sum y_i} - \frac{n+1}{n} \right] \frac{n}{n-1}$$

Where n is the total number of authors, and  $y_i$  is the number of journal articles published by author i, the authors being sorted in non-decreasing order of journal articles number.

### Web-Appendix 3

Publication types attached to each article and their co-occurrences, among all journals in the United States National Library of Medicine catalog having at least one Broad Subject term and having published at least 50 signed articles between 2015 and 2019.

Journal Article	0	12	77	10	160	205 156	33 847	4 582 473
Published Erratum	0	6	10	0	47	3	58 624	33 847
Case Reports	0	143	352	7	31 815	250 736	3	205 156
Letter	0	13	117	0	194 718	31 815	47	160
Introductory Journal Article	0	0	11 881	18 127	0	7	0	10
Editorial	0	13	127 348	11 881	117	352	10	77
News	0	27 853	13	0	13	143	6	12
None of these Publication Types	10 114	0	0	0	0	0	0	0
	None of these Publication Types	News	Editorial	Introductory Journal Article	Letter	Case Reports	Published Erratum	Journal Article

## Web-Appendix 4

Main characteristics of all journals in the United States National Library of Medicine catalog having at least one Broad Subject term and having published at least 50 signed articles between 2015 and 2019.

Year	All articles				
	2015	2016	2017	2018	2019
Number of journals with $\geq 50$ signed articles	4149	4143	4100	4065	4131
Number of articles					
Median [IQR]	130 [82 – 231]	131 [83 – 230]	132 [85 – 235]	136 [84.5 – 242]	142 [86.5 – 251]
Range	50 – 29 762	50 – 23 021	50 – 25 330	50 – 18 832	50 – 20 423
Number of articles with an author					
Median [IQR]	128 [82 – 228]	129 [82 – 227]	130 [83 – 233]	134 [84 – 240]	140 [86 – 249]
Range	50 – 28 937	50 – 22 827	50 – 25 327	50 – 18 635	50 – 20 423
PPMP (%)					
Median [IQR]	3.8 [2.4 – 5.8]	3.7 [2.4 – 5.8]	3.7 [2.4 – 5.7]	3.6 [2.3 – 5.6]	3.6 [2.3 – 5.5]
Range	0.2 – 39.2	0.1 – 46.3	0.2 – 50.4	0.1 – 52.1	0.08 – 44.4
95th percentile	11.4	10.8	10.4	10.1	10.4
Number of articles by MPA					
Median [IQR]	5 [4 – 8]	5 [3 – 8]	5 [3 – 8]	5 [4 – 8]	5 [4 – 8]
Range	1 – 195	1 – 172	1 – 194	1 – 161	1 – 226
More than one author tied as MPA	1 635 (39%)	1 623 (39%)	1 631 (40%)	1 655 (41%)	1 616 (39%)
Gini index					
Median [IQR]	0.087 [0.055 – 0.132]	0.086 [0.054 – 0.128]	0.085 [0.054 – 0.127]	0.085 [0.055 – 0.127]	0.085 [0.055 – 0.129]
Range	0.00 – 0.670	0.00 – 0.676	0.00 – 0.658	0.00 – 0.756	0.00 – 0.631
95th percentile	0.222	0.224	0.215	0.212	0.216
Median of publication lag ratio (MPA / no MPA)					
Median [IQR]	0.84 [0.55 – 1.12]	0.84 [0.57 – 1.12]	0.85 [0.56 – 1.10]	0.83 [0.55 – 1.13]	0.86 [0.59 – 1.14]
Range	0.00 – 35.00	0.00 – 79.00	0.00 – 28.5	0.00 – 24.75	0.00 – 92.50
Not calculable	2 420 (58%)	2 295 (55%)	2 204 (54%)	2 178 (54%)	2 243 (54%)

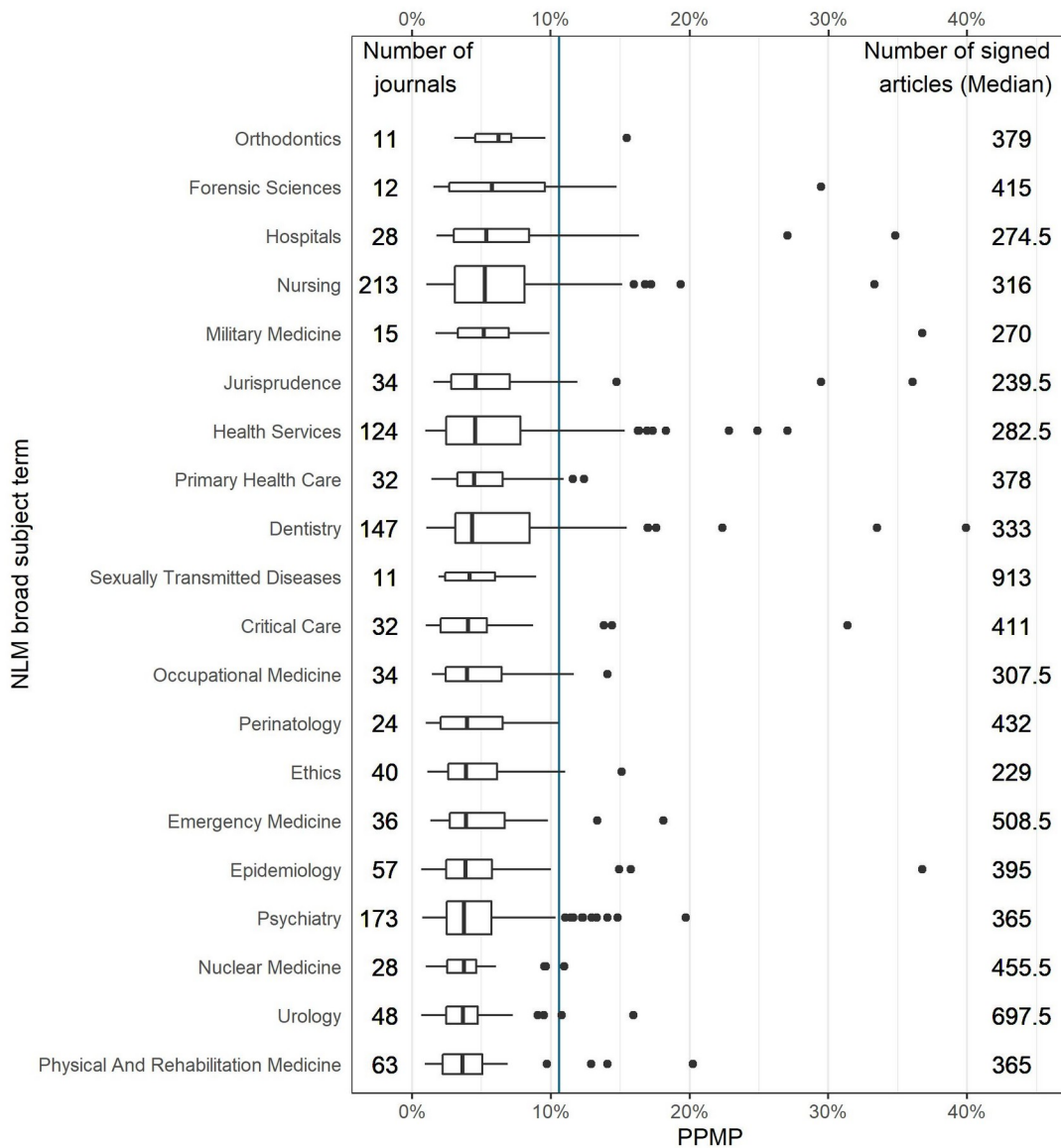
	<b>Journal Articles</b>				
<b>Year</b>	<b>2015</b>	<b>2016</b>	<b>2017</b>	<b>2018</b>	<b>2019</b>
Number of journals with $\geq 50$ signed articles	3994	4031	3977	3961	4020
Number of articles					
Median [IQR]	123 [80 – 214]	123 [80 – 213]	124 [81 – 221]	128 [81 – 226]	133 [83 – 232]
Range	50 – 28 108	50 – 22 142	50 – 25 322	50 – 18 832	50 – 19 877
Number of articles with an author					
Median [IQR]	122 [79 – 213]	121 [79 – 211]	122 [80 – 218]	127 [80 – 223]	132 [82 – 230]
Range	50 – 28 094	50 – 22 120	50 – 25 319	50 – 18 635	50 – 19 877
PPMP (%)					
Median [IQR]	3.5 [2.3 – 5.5]	3.5 [2.3 – 5.5]	3.5 [2.3 – 5.3]	3.5 [2.3 – 5.3]	3.4 [2.2 – 5.3]
Range	0.2 – 41.2	0.1 – 43.9	0.2 – 59.2	0.1 – 55.0	0.08 – 50.6
95th percentile	10.0	10.2	9.8	9.6	9.8
Number of articles by MPA					
Median [IQR]	5 [3 – 7]	5 [3 – 7]	5 [3 – 7]	5 [3 – 7]	5 [3 – 7]
Range	1 – 109	1 – 88	1 – 193	1 – 140	1 – 217
More than one author tied as MPA	1 774 (44%)	1 733 (43%)	1 804 (45%)	1 811 (46%)	1 754 (44%)
Gini index					
Median [IQR]	0.077 [0.047 – 0.119]	0.077 [0.048 – 0.117]	0.078 [0.048 – 0.117]	0.077 [0.050 – 0.116]	0.076 [0.047 – 0.115]
Range	0.00 – 0.667	0.00 – 0.595	0.00 – 0.644	0.00 – 0.759	0.00 – 0.635
95th percentile	0.210	0.213	0.202	0.201	0.203
Median of publication lag ratio (MPA / no MPA)					
Median [IQR]	0.85 [0.59 – 1.12]	0.87 [0.61 – 1.14]	0.86 [0.60 – 1.11]	0.86 [0.61 – 1.14]	0.89 [0.64 – 1.14]
Range	0.00 – 35.00	0.00 – 79.00	0.00 – 4.04	0.00 – 16.69	0.00 – 7.66
Not calculable	2 335 (59%)	2 271 (56%)	2 163 (54%)	2 163 (55%)	2 229 (55%)



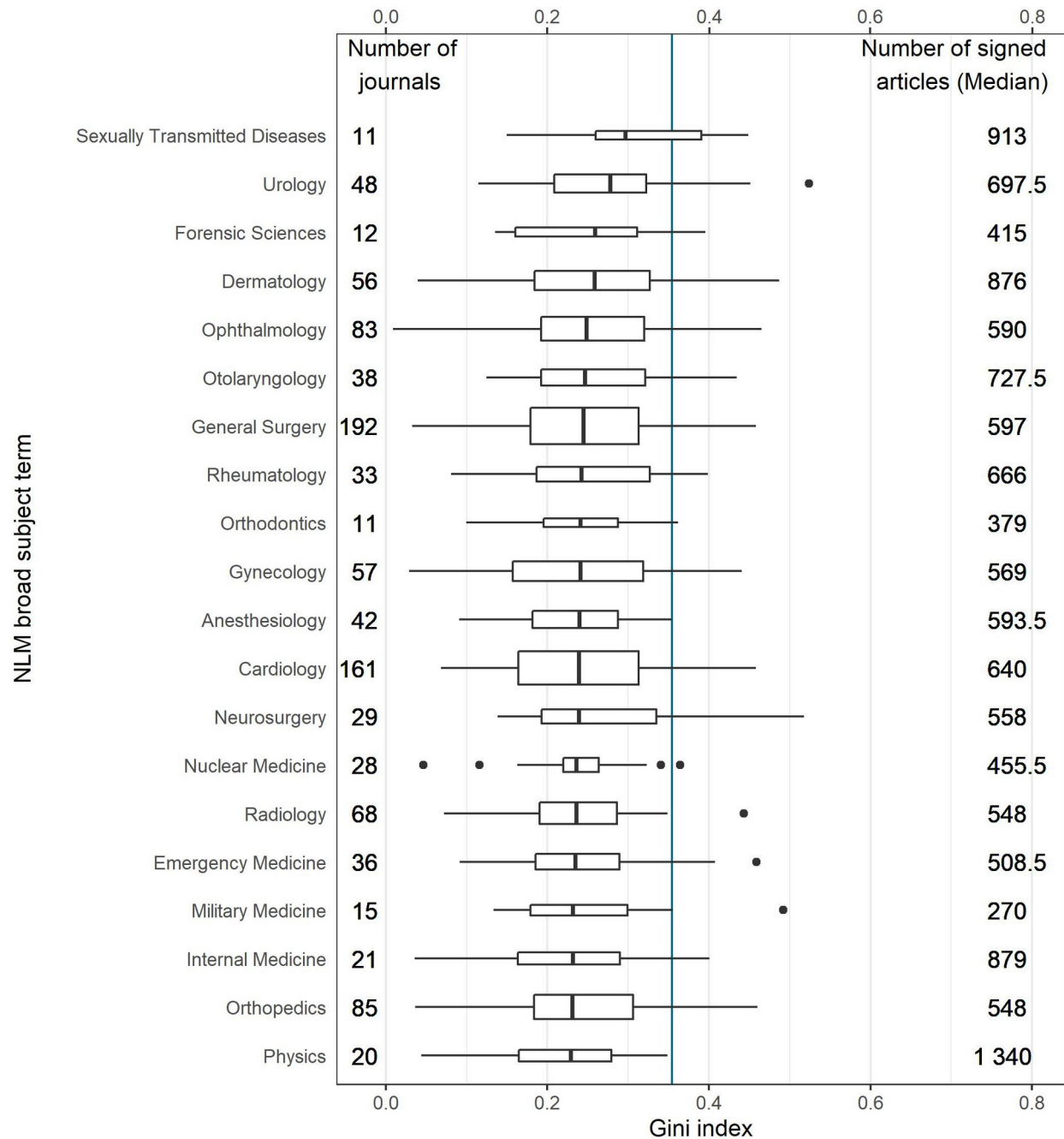
### Web-Appendix 5

Full appendix 5 is not reasonably printable. Here is a preview realized on the 20 fields with the highest median of PPMP or Gini index among its journals, among the fields represented by at least 10 journals:

Distribution of Percentage of Papers by the Most Prolific (PPMP) author among each United States National Library of Medicine (NLM) broad term, for journals having published at least 50 signed articles between 2015 and 2019. Width of the box-and-whisker are relative to the number of journals. Vertical line at the 95<sup>th</sup> percentile of PPMP among journals. Broad terms are arranged by decreasing median of PPMP.



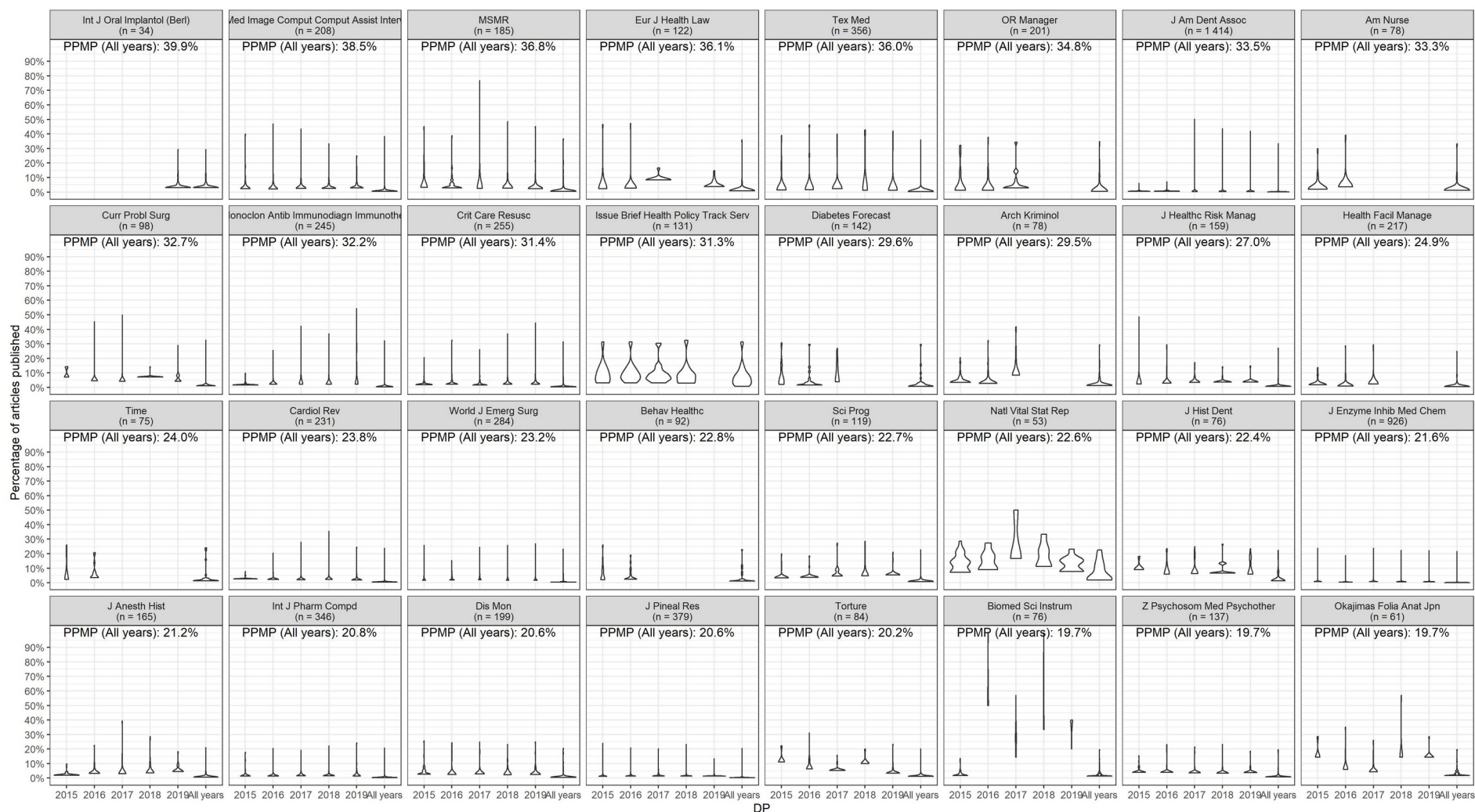
Distribution of Gini index author among each United States National Library of Medicine (NLM) broad term, for journals having published at least 50 signed articles between 2015 and 2019. Width of the box-and-whisker are relative to the number of journals. Vertical line at the 95<sup>th</sup> percentile of Gini index among journals. Broad terms are arranged by decreasing median of Gini index.



## Web-Appendix 6

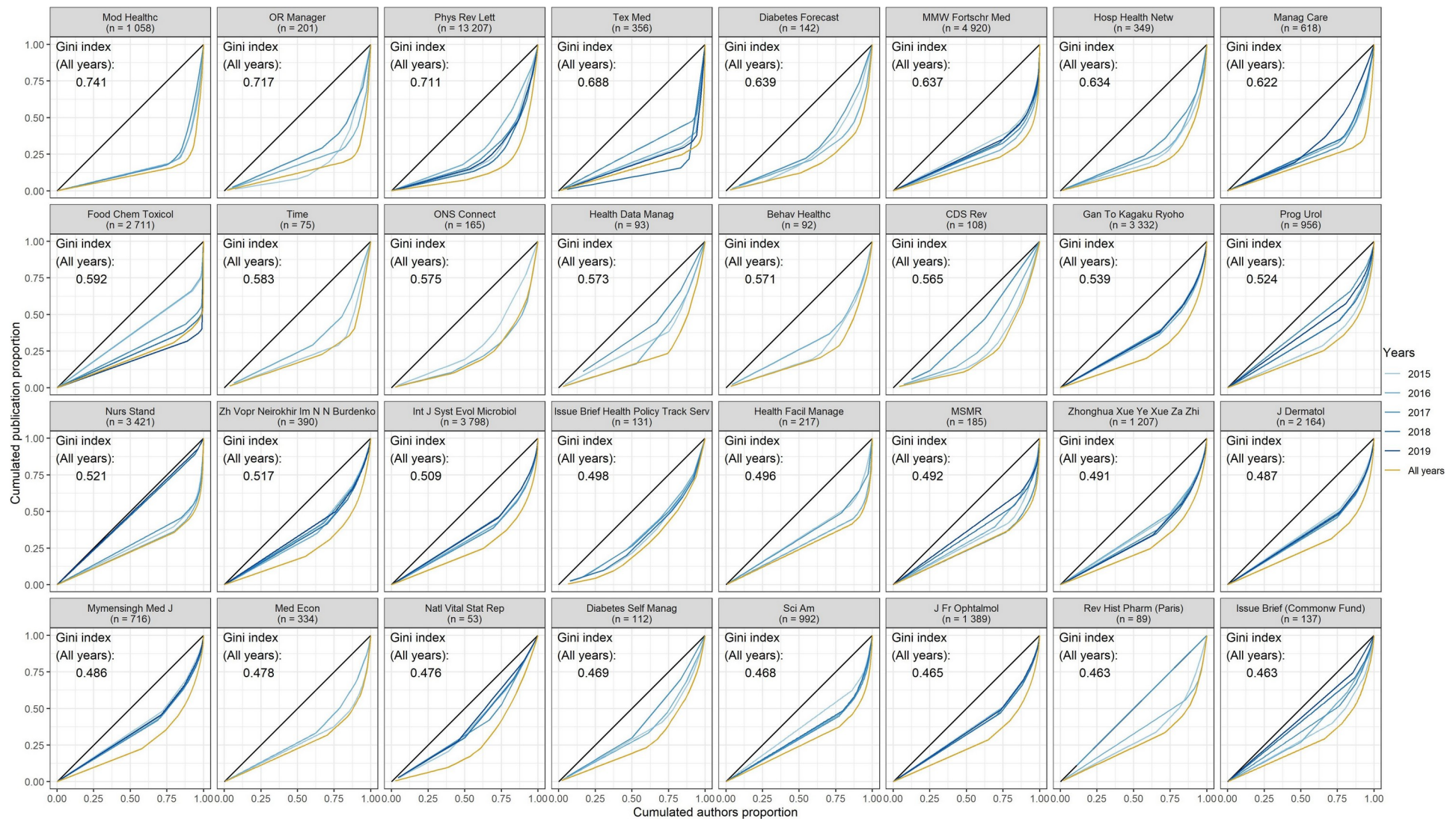
Full appendix 6 is not reasonably printable. Here is the first of 15 realised panels:

Yearly and global distribution of percentage of paper by each author for each outlier journal identified as having a Percentage of Papers by the Most Prolific (PPMP) author or a Gini index over the period from 2015 to 2019 of more than or equal to 95<sup>th</sup> percentile of journals having at least one NLM broad subject term. Facets are arranged by decreasing PPMP for the period 2015-2019.



Yearly and global Lorenz curve of distribution of appearances as an author, for each outlier journal identified as having a Percentage of Papers by the Most Prolific (PPMP) author or a Gini index over the period from 2015 to 2019 of more than or equal to 95<sup>th</sup> percentile of journals having at least one NLM broad subject term. Facets are arranged by decreasing Gini index for the period 2015-2019.

The Lorenz curve is a visual representation of inhomogeneity in distribution of authorship, with a straight line between 0 and 1 representing a perfect equality. The Gini index is the area between the line of equality and the Lorenz curve.



**SCANFF, Alexandre****Identification et caractérisation "revues d'auto-promotion" : une étude des journaux biomédicaux**

35 feuilles., 8 graphiques., 3 tableaux., 30 cm.- Thèse : Médecine ; Rennes 1; 2020 ; N° .

**Résumé français :**

**Context :** Le pourcentages d'articles par l'auteur le plus prolifique (PPMP) est une mesure simple pouvant être utilisée comme première alerte sur des journaux suspect de pratiques éditoriales biaisées. De même, une mesure plus complexe comme l'index de Gini a aussi été utilisée.

**Methods :** Pour tous les journaux biomédicaux, référencés dans la National Library of Medicine, avec au moins un Broad Subject Term attribué, et au moins 50 articles publiés avec les noms de leurs auteurs entre 2015 et 2019, nous avons calculé et décrit le PPMP et l'index de Gini.

**Discussion :** Nos résultats révèlent des relations entre auteurs et éditeurs qui interrogent sur l'indispensable indépendance des publications scientifiques. Les limites de cette étude incluent un risque d'incertitude sur les calculs des indices et une restriction aux journaux avec un Broad Subject Term attribué par la NLM qui peuvent avoir sélectionné des journaux plus établis.

**Résumé anglais :**

**Context:** The Percentage of Papers by the Most Prolific (PPMP) author is a simple measure that can be used as a red flag to identify journals that are suspected of biased editorial decision-making. Complementally, a more complex measure like the Gini index was used.

**Methods:** For all biomedical journals referenced in the National Library of Medicine, with any attributed Broad Subject Terms, and at least 50 papers with authors names published between 2015 and 2019, we calculated and describe the PPMP and Gini index of distribution of articles among authors.

**Discussion:** Our results reveal relationships between authors and editors that question the necessary independence of scientific publishing. Limits of this study included automated calculations with a risk of inaccuracy and a focus on journals indexed with NLM Broad Subject Terms that may have selected the more established journals.

Rubrique de classement :

Mots-clés : Bibliométrie, Pratiques Éditoriales, Publication, Intégrité Scientifique, Journalologie

Mots-clés anglais MeSH : Bibliometrics, Editorial Policies, Publishing, Scientific integrity, Journalology

Président : Pr Jean-François Viel

Pr Florian Naudet (Co-directeur de thèse)

JURY : Assesseurs :

Dr Clara Locher (Co-directeur de thèse)

Pr Matthieu Revest

Pr Ronan Garlantezec