NOVA IMS Information Management School

MDSAA

Master Degree Program in

Data Science and Advanced Analytics

Investigating the delay times in academic publishing

An empirical study on publishing delay times in academic journals

Oguz Kokes

Dissertation

presented as partial requirement for obtaining the Master Degree Program in Data Science and Advanced Analytics

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

NOVA Information Management School

Instituto Superior de Estatística e Gestão de Informação

Universidade Nova de Lisboa

INVESTIGATING THE DELAY TIMES IN ACADEMIC PUBLISHING

by

Oguz Kokes

Dissertation report presented as partial requirement for obtaining the Master's degree in Advanced Analytics, with a Specialization in Data Science

Supervisor / Co Supervisor: Bruno Miguel Pinto Damasio

November 2022

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration. I further declare that I have fully acknowledge the Rules of Conduct and Code of Honor from the NOVA Information Management School.

Oguz Kokes

November 29, 2022

ACKNOWLEDGEMENTS

I am forever grateful to Serdar Cetiner and Summani Duzenli for their continuous and limitless hardware and emotional support throughout the data collection process, patiently helping me with countless data collection pipelines and errors.

I also would like to thank Cevdet Yilmaz and Hakan Can for their ongoing and amazing hospitality and friendship, opening their homes and allowing me to work on my dissertation.

ABSTRACT

The aim of this study is to analyze the differences in article publishing delay times from different perspectives. Previous works on the topic suggest there are significant differences between article publication times, which has a direct effect on the author's personal and professional life. However, as the required dates for the analysis are not available in article databases, the works on the topic are limited to certain publishers/databases.

Using the entire Scimago Journal Ranking Q1 journal pool, this study creates a representative and comprehensive article dataset, containing submittance, acceptance and publication dates for over 200,000 sampled articles from 27 different subject areas between 2010-2020. This allows publishing delay times to be analyzed from different perspectives and offers a baseline for any future studies.

The study shows clear delay time differences between subject areas. The shortest delay time occurs in Life Sciences articles, with an average delay of 6 months, three times quicker overall than Social Sciences articles. Publication year analysis shows that while delay times are improving over time, this improvement is coming from acceptance to publication time delay, driven by the increase of digital publications. Delay times do not show the same improvement for the more problematic submission to acceptance delay, highlighting the reviewing process. Open Access journals offer an alternative to the traditional publications, and are faster overall, however their performances started to stagnate as number of publications increased each year.

Author affiliated country data is not balanced, and the dataset is dominated by submissions from certain countries, namely United States, China, United Kingdom, and Germany, indicating these countries' overall dominance on the scientific domain. However, matching analysis shows that an affiliated county's "Global North vs. South", "English as first language", and "G7 membership" status do not play a significant role in their subsequent delay times, indicating a fair refereeing.

KEYWORDS

Academic publication; Publishing delay; Publishing delay times;

INDEX

Contents

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

SJR	Scimago Journal Ranking
CR	CrossRef
ASJC	All Science Journal Classification
DOAJ	Directory of Open Access Journals
ISSN	International Standard Serial Number
DOI	Digital Object Identifier
OA	Open Access

Total delay time	total duration from an article's submission to publication
Acceptance delay time	duration between an article's submission to publisher'sa cceptance
Publication delay time	duration between an article's acceptance to its publication

1. INTRODUCTION

1.1. PUBLICATION PROCESS

Academic publications are the cornerstone of scientific advancement. Once a paper is published, the information uncovered and shared by its author is made available to its scientific community and contributes to the overall knowledge in their domain. The alternative, failure to publish, is described as a "scientific crime" by some, means years of work getting lost along with its researcher. This affects both current and future works and scientists, delaying the progress of their domain. (Clapham, 2005) Although significant, a paper's publication process can hardly be described as equally positive or inviting for the authors (Huisman & Smits, 2017).

The first step, submitting the paper to a journal, is also where most authors get stuck, as a study containing 2,371 academic journals from 2017 shows that the average acceptance rate to be 32% percent (Herbert, 2019)**,** meaning there is always a high possibility of a journal straight-out rejecting your paper, known as "desk rejection", sometimes after months of waiting for a response (Björk, 2019). As most publishers demand exclusive submissions, each rejection means starting over; choosing a new journal to submit to and waiting for an answer from.

A positive answer is also far from the final product, as before a paper is officially accepted by the journal, authors and publishers, tend to enter a back-and-forth review & revision stage (Björk, 2019), known as peer-reviewing, where the paper is judged for its scientific validity, topic and formatting among other aspects. These repetitive revisions can take months of author's time, requiring numerous and major changes to the original work and causing huge amount of extra work and frustration (Powell, 2016), and eventually pushing back the paper's publication time. However, this is not to argue that the peer-review is an unnecessary step in publication, only there for annoyance of the authors. Nicholas et. al explain that despite its negatives and shortcomings, their study with 4,000 academic researchers shows that peer-reviewing is still viewed as the "central pillar of trust" by the academic community, accepted as a crucial step during publication to ensure the validity and quality of work, in a way a necessary gatekeeping to set and maintain the high standards of academia (Nicholas, et al., 2015)**.** Siler et al. support this claim in their paper, where they conclude that peer-reviewing is in general a necessary and valuable part of the publication process to maintain a certain quality, and peer-reviewers and editors are generally good at recognizing eventually highly cited papers as they analyze that papers that received lower scores in peer-reviews ended up with lower citations after their publications (Siler, Lee, & Bero, 2015)**.**

Moving on with the process, after the peer-review is complete and the revisions are approved by the reviewers, the paper is finally officially accepted to be published by the publisher, and it is put into the journal pipeline. However, its publication, which is now completely out of the hands of the author, is a time demanding process, especially for traditional paper publications (Larivière, Haustein, & Mongeon, 2015).

1.2. "PUBLISH OR PERISH" AND "PRIORITY RULE"

Going back to Clapham's "scientific crime" argument and looking at the publication process, aspect of authors' motivation, or why do academic researchers go through this process can be questioned. From a holistic "big picture" perspective Clapham's approach can be accepted, but it would be naïve to

assume an author's sole motivation as scientific contribution. A more understandable approach can be "publish or perish", a principle which states that a researcher's academic success is directly dependent on their number of publications (De Rond & Miller, 2005). As hiring, promotion, tenure, and academic prestige are all based on their article count, this creates an environment where a researcher must deliver continuous publications if they want to be successful in their fields (Neill, 2008). With limited platforms to publish, publication process becomes a deciding factor in a researcher's academic career and creates an overall competitive environment.

"Priority rule" adds another layer to the competitiveness argument, as it states that getting published alone is not enough. Being first to do so and taking the main credit for a finding is just as or even more crucial for a successful academic career (Hill & Stein, 2021). As Strevens explains it, scientific discovery is a "winner takes all" race, and there are only prizes for first place (Strevens, 2003), meaning researchers not only have to get published, but they must do so before their fellow colleagues, being the one responsible for advancing their domain further, achieving personal and professional glory (Reif, 1961).

1.3. PUBLISHERS

This brings the discussion to the other protagonists of the publishing domain, the publishers. The publication process clearly shows that researchers cannot be the decider of their own fate alone, regardless of their amount of research and quality of submission. Publishers are eventually the judge, jury, and executioners of this domain. Ultimately, this is not a race for who submits their manuscript first, but rather who first gets published, and by deciding which submission gets published and when, publishers also elect the winners (and the losers) of the academic race.

Academic publishing is not a new term with unknown institutions; however, with thousands of journals and publishers, with even more expected in the future, it does require further exploration (Taskin, Taskin, Dogan, & Kulczycki , 2022). In his paper Nishikawa-Pacher combines four main article databases to list the top 100 publishers in the world in terms of journal count. The paper points out the top 5 publishers (Springer, Taylor & Francis, Elsevier, Wiley and SAGE), called as the "the usual oligopoly of major publishing" in the paper, are the only publishers to have more than 1,000 journals to their name (Nishikawa-Pacher, 2022). Another interesting note is that out of the 100, only 17 publishers belong to the Global South University Press, highlighting the overwhelming Global North representation in terms of academic publishing.

Nishikawa-Pacher's paper is not the only one to use the "oligopoly" claim, as Larivière et al. conduct an analysis of published articles between 1973-2013 to arrive at similar results. They indicate that despite popular belief, these top 5 publishers have in fact increased their coverage during the "digital era", starting from the 1990s, and are responsible for more than 50% of the total papers published in 2013 (Larivière, Haustein, & Mongeon, 2015).They conclude that this naturally caused an increase in these institutions' profit margins and created a dependency from the researchers.

1.4. BIBLIOMETRIC RESEARCH & PREVIOUS WORKS

As the race for publishing and publisher preferences determine the fate of researchers and their domains, researchers in return have naturally been interested in analyzing and understanding the process in detail. The overall amount of time spent (or wasted, according to some) until a papers

submission to its publication has been defined by the numerous papers on the issue as “publication delay”, “publishing lag”, “editorial delay” among other names, and have been the topic of many bibliometric researches (Taskin, Taskin, Dogan, & Kulczycki , 2022).

Björk and Solomon (Björk & Solomon, 2013) conducted one of the first meta-analysis on the topic, defining the publishing landscape across different categories/viewpoints. In their paper, they put a single year in focus (2012) and analyzed 2700 articles from 135 journals from all 27 Scopus subject areas. Their work points out that publication delay times show a very clear variation from subject area to subject area, with business & economics articles, the “slowest” subject area with just under 18 months, have twice as much as publication delay than chemistry related articles with a delay of 9 months. Open Access (OA) status of a journal also plays a significant role as they report that OA journals have considerably shorter publication times compared to subscription-based counterparts.

A more recent meta-study by Himmelstein (Himmelstein & Powell, 2016), examines around 3 million articles covering 20 years, using data exclusively collected from PubMed database from 1997 to 2016. Himmelstein indicates that although the overall publication times have been getting shorter over time with 69% of journals decreasing their publication delay times, the median time for acceptance has been around 100 days for more than 30 years, as acceptance times have been decreasing for some journals while increasing for others. Lee et al. (Lee, Kim, & Lee, 2017) also support this claim, as they argue that the improvements in publication delay are caused by the decrease of what they define as “lead lag” or acceptance to publication times, rather than peer-reviewing process. Asaad et al. (Asaad, Rajesh, Banuelos, Vyas, & Tran, 2020) studied the publication delay in plastic surgery journals, to find out very similar results, the decrease is there but it is coming from the time between acceptance and publication, pointing out the reviewing process as the main culprit for delays. Building upon and working together with Himmelstein in a follow-up study, Powell indicates that PLoS ONE, a popular open access journal family, with the motto “Discover a faster, simpler path to publishing in a high-quality journal.” aiming to simplify the reviewing process, doubled their review times in the last 10 years, from 50–130 days to 150–250 days (Powell, 2016).

With all the papers so far singling out the reviewing process and acceptance delays, we need to analyze this step in more detail to better understand why peer-review delays are not improving along with overall publication times. Firstly, there are simply more authors submitting their manuscripts than ever before (Lyman, 2013). Even Huisman et al., who approach the issue from the author’s perspective, state that reviewers are overloaded with review work and rarely receive any compensation or acknowledgement for their work in return. This causes a reviewer, who is also an academic researcher with their own submissions and research to deal with, to give less priority to their reviewing tasks (Huisman & Smits, 2017). This little “concern” from the reviewers’ side, combined with the increasing number of submissions, can be one of the explanations behind acceptance delay.

Mrowinski et al. (Mrowinski, Fronczak, & Fronczak, 2020) have a different take on the issue and approach the peer-reviewing from journal editors’ point of view. Their paper argues that journals are “flooded” with external submissions, which rarely follow journal guidelines, causing the local and “technically correct” submissions to be accepted quicker and more frequently. The external authors also tend to re-submit their paper without the necessary technical or scientific corrections, inflating the number of submitted papers, decreasing the acceptance rate, and causing extra work for the editors. However, they also note that the editorial responses, whether it being acceptance or rejection,

are twice as fast for local submissions than external ones and local articles are more likely to be accepted for publication.

This local versus external authors issue introduces us to the highly discussed topic of "bias" in academic publishing. Zhu defines this specific issue as "home country bias" in his paper where he argues that articles with at least one US based author make up majority of the publications of *Journal of Medicine* between 2000-2019 while their subsequent citations have been lower than non-US articles (Zhu, 2021). van Lent et al. give more proof and context in favour of this argument, as their work investigating submissions to eight medical journals conclude that European (25.4%) and US (26.8%) based authors together have a significantly higher publication ratio for their submissions compared to authors from the rest of the world (12.4%) (van Lent, Overbeke, & Jan Out, 2014). Adding a more optimistic outlook to Zhu's findings, Hsiehchen et al. suggest that the US dominance in medical publishing is decreasing over time, as the US based author affiliation in clinical imaging decreased from 55% to 29% comparing two 5-year intervals starting from 2000 and 2010 (Hsiehchen & Espinoza, 2016).

All this evidence shows that different approaches to the publication delay times can bring different explanations to the overall issue, whether being more circumstance (subject area, journal, year, etc.) or author related. However, one thing is for certain, no matter how you approach the issue, it is hindering scientific advancement, negatively affecting the authors both on a personal and professional level and slowing down the advancement of their domains.

Therefore, the purpose of this study is to create an all-encompassing snapshot of the publishing domain covering a 11-year period between 2010 to 2020 without being limited to a specific subject area, journal, publisher, or database, creating an article publishing times dataset to approach and answer the question of ***"How do publication speeds differ?"*** from the perspective of their domain and their authors.

2. METHODOLOGY

2.1. SAMPLING DESIGN

2.1.1. Feasibility Study - Retrieving Article Submitted & Accepted Dates

The first and arguably the most crucial step of this project is to solve the issue of retrieving article submitted and accepted dates. Except for PubMed, which includes the dates as a part of article, and Elsevier, where the dates are not a part of the article metadata but are accessible via the full text of the article, submission and acceptance times are generally either not a part of the metadata fields, incomplete, or not shared by the publishers in other academic databases.

Moving one step further, the article itself may or may not contain this information, entirely depending on the journal's editorial & design choices, as some journals have included this information in certain year's publications but not for others. Journals that choose to include the dates also have greatly differing formats for disclosing this information, some only including submission or acceptance dates, some including both, and some also including revised submission and acceptance dates. These dates can be placed after the title of the article, after the abstract, as a footnote, or before the References section; with exact (DD/MM/YY) or only month (MM/YY) given.

Considering these issues, the sampling design must include a feasibility study to check which journals are accessible via university sources and eligible for this study, allowing us to extract the necessary dates (both submitted and accepted date) directly from article texts (via their PDFs). This way, the data collection process will be timewise and computationally consuming, but it can be independent of any single data source and can contain all eligible journals. The extracted information can then be utilized along with the rest of its metadata.

2.1.2. Data Sources

The framework below was used to check the eligibility of journals and to create the necessary three datasets for the project, one for journals, articles, and authors.

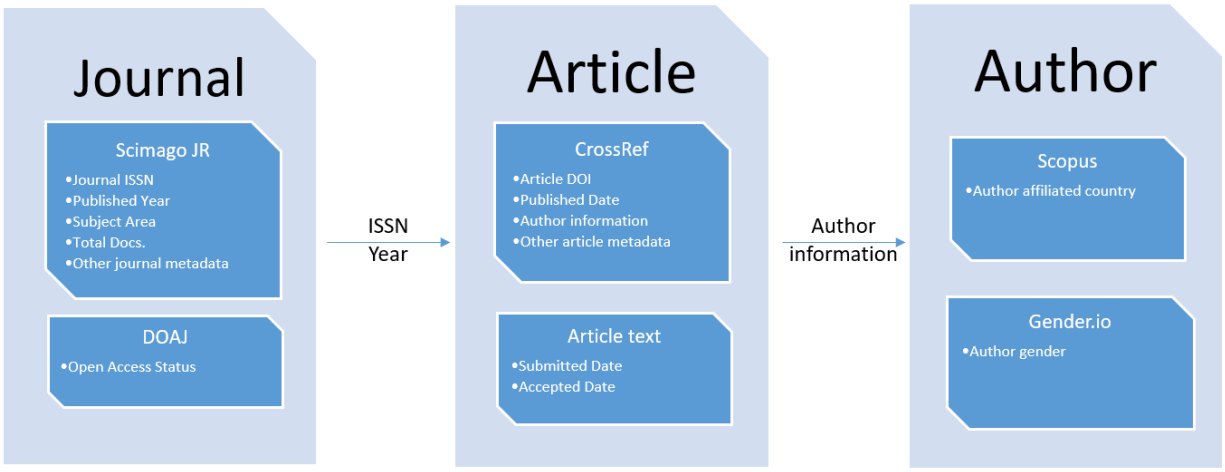


Figure 1: Data collection framework of the project. The figure displays the three datasets created for this study, including their sources and which fields were retrieved from there

2.1.2.1. Journals

Scimago Journal Ranking (SJR) was used to extract journal related information and as a base point for this project. This is because the institution includes a large variety of journals spanning over the targeted years, includes different publishers, and covers all 27 “Subject Areas” and their 333 “Subject Area Classifications” defined by All Science Journal Classification (ASJC) System used by Scopus. It also offers an inherent journal ranking, SJR (Scimago Journal Rank) score, which is used to categorize journals into 4 quartiles on a subject area classification level. It also includes the total number of articles published by a journal each year, allowing us to calculate the total population of publications each year vs. subject area.

The bibliometric data related to all journals available in the Scimago Journal Ranking portal between 2010 – 2020 were collected from their official website. A total of 17 metadata fields were retrieved for each journal record. Each ISSN & Publication Year pairing was considered as a unique journal record (and given a unique JRNL_ID). Only the journal records within the first quartile (Q1) were considered. This amounted to 75,530 journal records, which amounts to 12,563,133 articles.

Directory of Open Access Journals (DOAJ) was used to retrieve the Open Access status of journals, using the year journal went Open Access.

2.1.2.2. Articles

CrossRef was used as the article metadata source of the project, as encouraged by the Initiative for Open Citations for bibliometric analysis (Hendricks, Tkaczyk, Lin, & Feeney, 2020). CrossRef official website indicates that as of 2022, it includes 134M metadata records, making it a comprehensive and suitable source for retrieving article metadata for this study. Its highly effective Rest API, as mentioned by Hendricks, allows to query journal ISSN and publication year, readily available from Scimago, to get article Document Object Identifier (DOI) and other relevant metadata, effectively functioning as a search engine and converting journal information into article metadata.

As not all journals are available on CrossRef (CR), journal records were filtered based their availability on the CrossRef database. Total number of available articles were retrieved for each journal record using CrossRef Rest API. Any unavailable (CR Article Count equalling to 0) record was marked as ineligible, decreasing the number of eligible journal records to 64,218. Article count was filtered using the lower between the Total Docs. field of Scimago and CR Article Count was used, making the eligible article count 10,229,213.

Each journal record available on CrossRef were then checked if their texts or PDFs are available through one of the available university sources using a custom date retrieving pipeline. For each PDF accessible journal record, 3 articles were randomly selected and checked whether the article contains the necessary (submitted and accepted) dates. If available, PDFs’ text was retrieved by streaming the file contents from the source. As retrieved texts can be in very different formats, it was then processed through an NLP pre-processing pipeline to standardize before text extraction. Finally, the text is scanned with a regex pattern for any matches with both “submitted date” and “accepted date” information. The pattern checks for a combination of different string, date and punction formats to account for journal stylistic differences. If both fields have a positive match, the matched texts are converted into datetime format and returned.

If no dates were retrieved from any of the 3 trial articles, that record was marked as not ineligible. Successful trials were saved to be later used in the article dataset. The final number of eligible journal records and articles can be seen in the table below. It should be iterated that these numbers are theoretical and may not represent the actual number of articles with date information given.

	Scimago	Available on CrossRef	Both submitted and accepted date given
Journals	75,530	64,218	27,597
Articles	12,563,133	10,229,213	7,158,605

Table 1: Number of eligible Q1 journal records and articles.

2.1.2.3. Authors

The metadata retrieved from CrossRef generally includes basic author information, however, author affiliation is not included for most articles. Therefore, a separate source was necessary to retrieve this information. Scopus database allowed utilizing the already available name, surname, ORCiD ID and related subject area information to be queried through its AuthorSearch API, enabling the collection of author affiliation and its subsequent affiliated country information. As author data collection was done after article dataset is completed, any articles with missing author information were removed retrospectively.

Once the author affiliated country information was retrieved it was then mapped using pycountry, a-world-of-countries and country-converter Python packages, allowing the separation of countries as "Global North vs. Global South", "Native English speaking vs. non-native English speaking" and "G7 countries vs the rest of the world ". It should be noted that as European Union is counted as the "non-enumerated member", EU member countries are also included in this grouping.

2.2. DATA COLLECTION

The feasibility study shows that with the current framework around 42% of the journals' submitted and accepted dates are accessible. To calculate the necessary sample size for a representative article dataset, each journal's (Total Docs.) article count was allocated between its subject areas, creating a subject area vs year matrix. Each journal's SJR "subject area classifications" were mapped to their main 27 subject areas for each journal. Any remainders after the division were randomly allocated between the subject areas. The resulting matrix can be seen in the Appendix.

As an example, Journal of Finance published 66 articles in 2018, according to SJR. Checking its 3 subject area classifications; "Accounting" belongs to "Business, Management and Accounting", whereas "Finance" and "Economics and Econometrics" belong to "Economics, Econometrics and Finance" subject area. The articles were then divided between them as 22 and 44 respectively.

Using the population matrices, the required sample size was calculated for each cell, for margin of error of 4% at the 95% confidence. The resulting required sample size matrix can be seen in the

Appendix. The total number is 171,936 articles. Comparing this number with the eligible number of articles, a representative dataset could be created.

The process is repeated for journal counts, to ensure maximum journal representativeness. Due to the data collection pipeline, each eligible journal has at least one article contributing to the article dataset (which were already retrieved during date checking step). However, it is seen that for some subject areas and years, the eligible journal count is less than the required journal sample size.

Moving on to the article data collection step, eligible journal records were filtered for each cell of the required sample size matrix, based on their publication year and subject area. The required article sample size was then randomly distributed among the filtered journal records, not exceeding their total document or CrossRef article counts.

Journal records were queried via CrossRef Rest API, using ISSN, publication year and the required article count, using the API's sampling function to get article DOI and 11 other selected metadata fields. As it is impossible to identify the specific subject area for each article, each was assigned to a subject area depending on the journal record. CrossRef metadata contains two publicaton fields (published-print and published-online), and the earlier date was used as the article's official publication date, as it is the date an article is officially available. If only month and year was given for a publication, 15th of that month was asssumed. For successfully metadata retrieved articles, submitted & accepted date information was extracted via the date extraction pipeline. Extracted dates were used to calculate the three delay periods: total delay time, submitted to accepted delay (acceptance delay) and accepted to publication delay (publication delay). To compansate for any possible errors, required article count was exceed for every journal record if possible. The resulting dataset can be viewed in the Appendix.

It should be noted that there are insufficient articles for certain matrix cells. This is due to the low number of eligible journal records, and thus articles, for those subject areas. These are *Multidisciplinary* (2010: 198, 2011: 252, 2012: 25) and *Dentistry* (2011: 4, 2013: 1, 2014: 23, 2015:3) with a total of 475 and 31 missing articles respectively.

Scopus AuthorSearch API was used for retrieving author affiliated country information, where only the first authors were taken into consideration. If an article had muliple first authors, the highest represented country was selected. If the numbers are equal, the "first" first auhtor's affiliated country was chosen. Any articles with missing or empty author affiliation fields were removed from the article dataset. These retrospective changes are reflected in the tables above. The final representation of the datasets can be seen in the figure below.

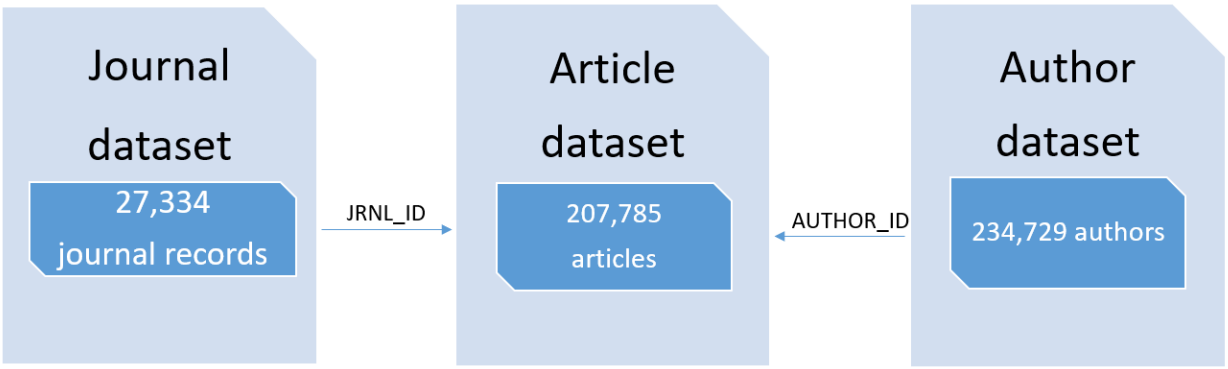


Figure 2: The final representation of the project datasets, indicating the total number of records in each dataset and their relationship to each other

2.3. Data Analysis

Welsh's t-test and ANOVA were performed using SciPy Python library to test the significance of the findings during the descriptive analysis section. The dataset was split based on Subject Area, Published Year and Open Access status respectively, using total delay time, acceptance delay and publication delay as the analyzed parameters.

For the main analysis, the author affiliated country was analyzed using three separate perspectives (treatments) to split the dataset. These were selected as "Global North vs. Global South", "Countries with English as a first language vs. rest of the world" and "G7 member vs. rest of the world". They were selected as they can offer different geographical, cultural, economic, and language-based distinctions to split the articles. affiliated country. G7 member countries consist of Canada, France, Germany, Italy, Japan, the United Kingdom and the United States as well as all European Union (EU) countries, as EU is accepted as the "non-enumerated" member of the forum. For the "English as a first language" grouping, a country was marked as "native speaker" if English is the first official language of the country or is spoken by most of the population. It should be noted that the affiliated country's language and the authors' language are not necessarily the same feature, as many academics may choose to continue their studies abroad. However, the main approach here is that a native English speaking affiliated country may have more resources and opportunities to allow an author to improve their paper's linguistic quality. As the entire dataset is consisted of English written articles, we can use this aspect to see if being affiliated from such a country creates a difference in terms of publication delay times.

The affiliated countries can also be grouped and compared based on their G7 membership. G7 member countries consist of Canada, France, Germany, Italy, Japan, the United Kingdom and the United States as well as all European Union (EU) countries, as EU generally is accepted as the "non-enumerated" member of the forum. This allows a more specific approach and analysis to these countries, as most are also a part of the "Global North" grouping analyzed earlier.

Propensity score matching was used during this section to observe the differences between the affiliated country groupings in more detail. The fields analyzed in the first section (Subject Area, Published Year, Open Access Status) were used as covariates. "Publisher Origin" was also included as an additional covariate field, to account for any underlying publisher related variances. As dataset contains publishers from 47 countries with varying publication counts, they were grouped based on a

publication count threshold to represent this information effectively and practically. The field contains origin country of the publisher, if it is one of the top 4 (more than 10,000 total publications) publishing countries (United Kingdom, United States, Netherlands, and Germany), or the publisher origin continent (Europe, Asia, South America, Africa, Oceania and North America) instead. It should be noted that Oceania was preferred (rather than Australia) to account for any articles originated from New Zealand.

Due to the nature of the dataset, propensity score calculation and nearest neighbor matching allowed exact matching between articles for the selected covariates. The minority treatment group was used to perform one-to-two matching with replacement.

3. RESULTS AND DISCUSSION

Our research question "How do publication speeds differ?" can be approached and answered from different perspectives as also seen in the previous works on the topic mentioned earlier. While Solomon, Himmelstein and several others take a more descriptive holistic approach, trying to define the rules of the publishing domain from subject area, year, journal accessibility perspectives, other authors such as van Lent have a more personal and author-oriented take on the issue, trying to analyze whether authors' inherent and mostly beyond their control characteristics play a part in their papers' publication delay times, as well as their own career growth.

I believe both approaches are necessary to fully understand academic publishing. Describing the rules of the domain can help us better comment on the deep dive analysis later, especially regarding a controversial topic such as bias. Simply approaching the issue directly from a "bias" perspective might cause us to make incorrect assumptions and come up with conclusions when there are none. In his paper, Gerring states that "Causality and description are intimately related; one cannot be understood without the other." (Gerring, 2012). Agreeing with Gerring's statement, I decided to split the analysis into two parts. The first part is following the holistic descriptive approach to define the main rules and patterns of the domain. This will also give us a unique opportunity to compare our findings with those of Solomon et al. and Himmelstein, as our dataset is not bound by a single year or data source. The second part can then approach the issue of bias, specifically from the perspective of author affiliated country and gender, using propensity score matching to analyze if given equal circumstances, there are conclusive differences among authors due to some underlying "unspoken rules" within academic publishing.

3.1. DESCRIPTIVE ANALYSIS

The first approach to describing and understanding our data can be following the subject area versus year matrix approach we used to collect the dataset in the first place. The mean delay times matrix table can be seen in Table 3.

A quick look at the table allows us to make several general comments on our dataset. From a subject area perspective, it shows "*Decision Sciences", "Economics, Econometrics and Finance"* and *"Business, Management and Accounting"* all have longer mean delay times throughout the years, varying between 500-600 days. Comparing these numbers to "*Immunology and Microbiology* and *Biochemistry", "Genetics and Molecular Biology",* which have total delay times under 200 days (except for a single year), shows there is a great disparity and difference across domains. While authors from certain subject areas must wait around 20 months to get their work published, other authors can theoretically publish three papers in the same time span. A yearly approach also displays that average mean time has decreased for all the subject areas over time. Although the improvement rates are different, all 27 subject areas had shorter delay times in 2020 than they did in 2010 indicating an overall improvement.

3.1.1. Subject Areas

Building on our observations above, the 27 subject areas can be mapped to their "main subject areas" (*Health Sciences*, *Life Sciences*, *Multidisciplinary*, *Physical Sciences* and *Social Sciences*) for a more compact view. Table 2 below shows the summary statistics for the main subject areas and indicates

the total time from submission to publication varies significantly ($p < 0.05$) across main subject areas. Life Sciences, which also includes all the "fast" subject areas mentioned above, is the fastest overall with a mean delay of 193 days. Social Sciences on the other hand, which contains all the "slow" subject areas above, is almost 150 days slower than its closest "competitor", which equals to 5 months of additional delay.

Main Subject Areas	Median Total Delay Time	Mean Total Delay Time
Life Sciences	166	193
Multidisciplinary	193	265
Health Sciences	208	243
Physical Sciences	241	287
Social Sciences	371	430
Overall	**237**	**292**

Table 2: Mean and median total delay times in days by main subject areas.

Table 4 presents average submission to acceptance and acceptance to publication delay times in days for the main subject areas. It appears that the same performance difference is also visible when times are observed separately. Social Sciences articles not only suffer from long acceptance times, but they also have the longest publication delay times, explaining their wide delay margin. Table also shows that Health Sciences has the 3rd fastest acceptance time, but their average 86 days of publication delay improves their overall ranking, making them the overall 2nd fastest publishing domain on average.

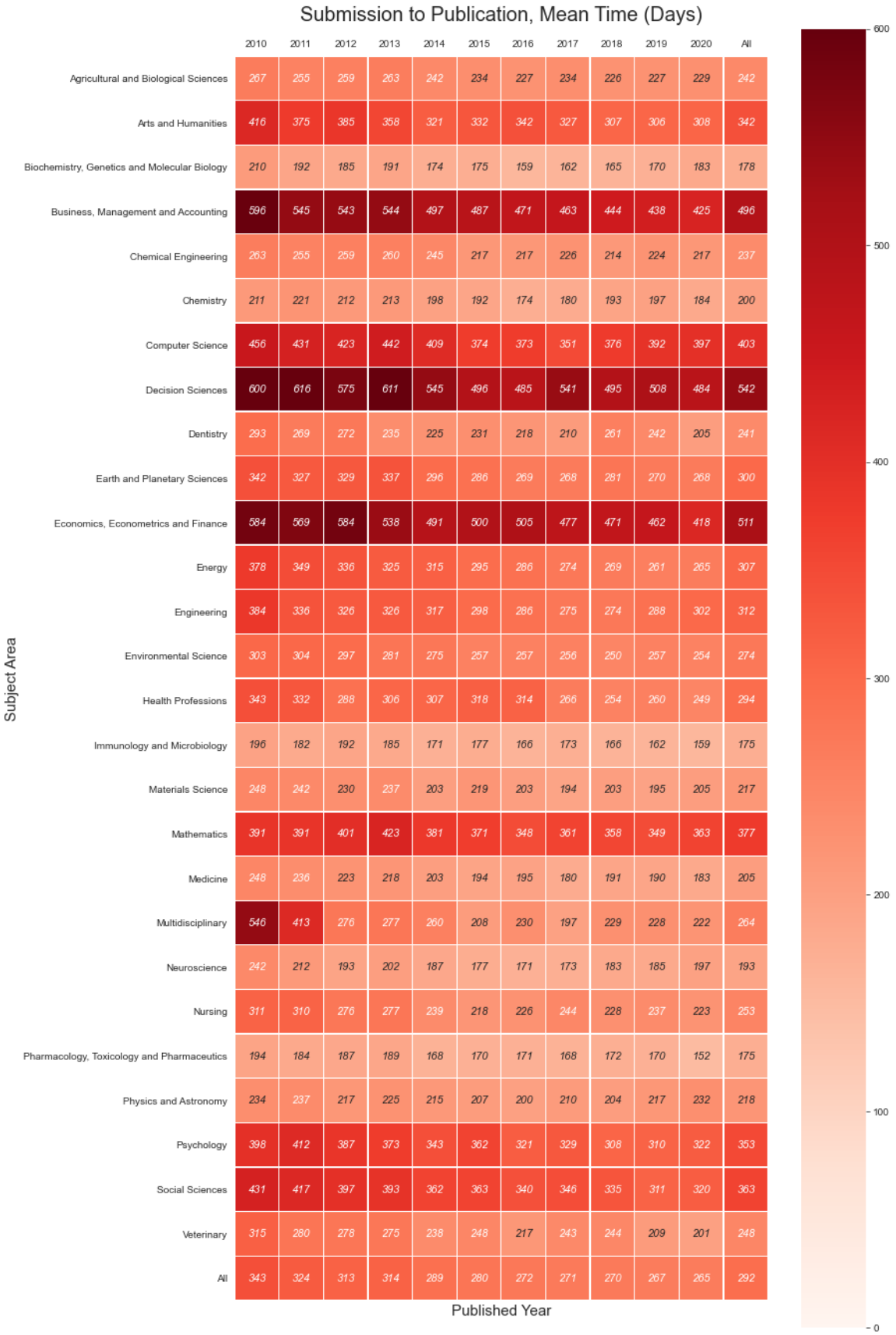


Table 3: Mean delay times table, indicating considerable differences between delay times of subject areas and years, with Decision Sciences in 2011 having the longest mean delay time with 616 days, and Pharmacology, Toxicology and Pharmaceutics in 2020 having the shortest mean delay time with 152 days.

	Life Sciences	Multidisciplinary	Health Sciences	Physical Sciences	Social Sciences
Submittance to Acceptance	125	151	156	181	292
Acceptance to Publishing	68	113	86	105	137

Table 4: Submission to acceptance and acceptance to publication delay times in days for the main subject areas

Moving on to subject area analysis, Figure 4 displays the delay times and their submittance to acceptance and acceptance to publication delay times breakdown. An interesting observation here is that the top categories are dominated by Life Science and Physical Sciences related subject areas, despite Physical Science being the 4th fastest main subject area on average. We can see that some physical sciences such as *Chemistry*, *Materials Science*, *Physics and Astronomy* have both short acceptance and publication delay times, but these are hindered by other slow subject areas belonging to Physical Sciences such as *Mathematics* or *Computer Science,* decreasing their overall ranking. We can also see that the bottom three subject areas have such long acceptance delay times that 19 of the 24 remaining subject areas have shorter overall delay times.

This can be further analyzed in Figure 3, displaying the subject area delay times boxplots from submission to acceptance. The top 8 shortest acceptance times are all from Life Science and Physical Sciences related subject areas, whereas all Social Sciences areas are included in the bottom 8, with Multidisciplinary and Health Sciences subject areas spread in between.

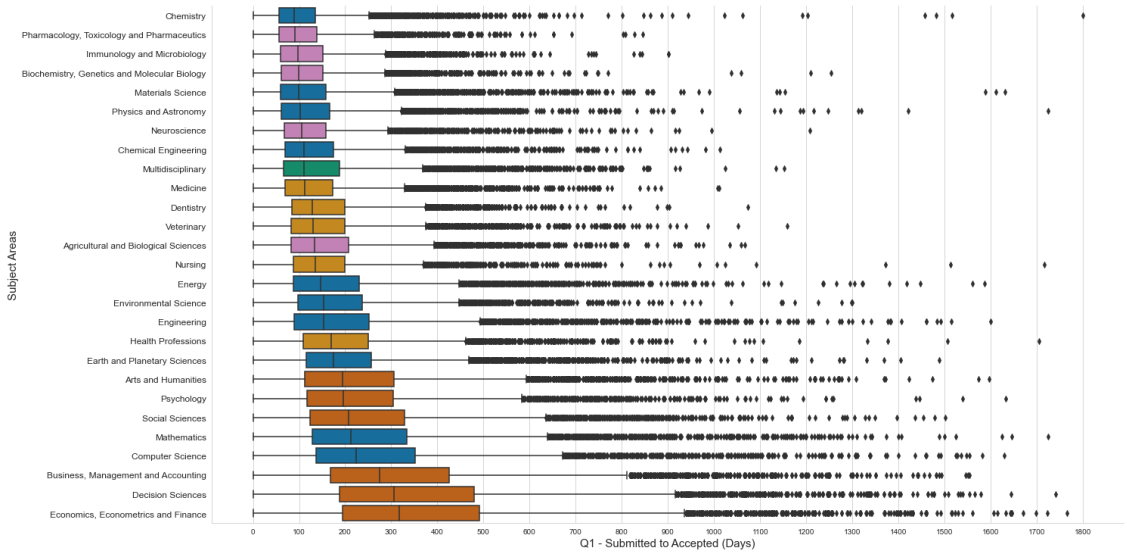


Figure 3: Publication delay times in days across subject areas, displaying the short acceptance times for most Health and Life Sciences articles, as opposed to long delay times of Social Sciences

Figure 4: Average delay times for subject areas, displaying both acceptance and publication delay times.

Figure 5 also displays the clear distinction between publication delays, Life and Health Sciences subject areas exclusively dominating the top of the graph, leaving all Physical and Social Sciences subject areas at the bottom.

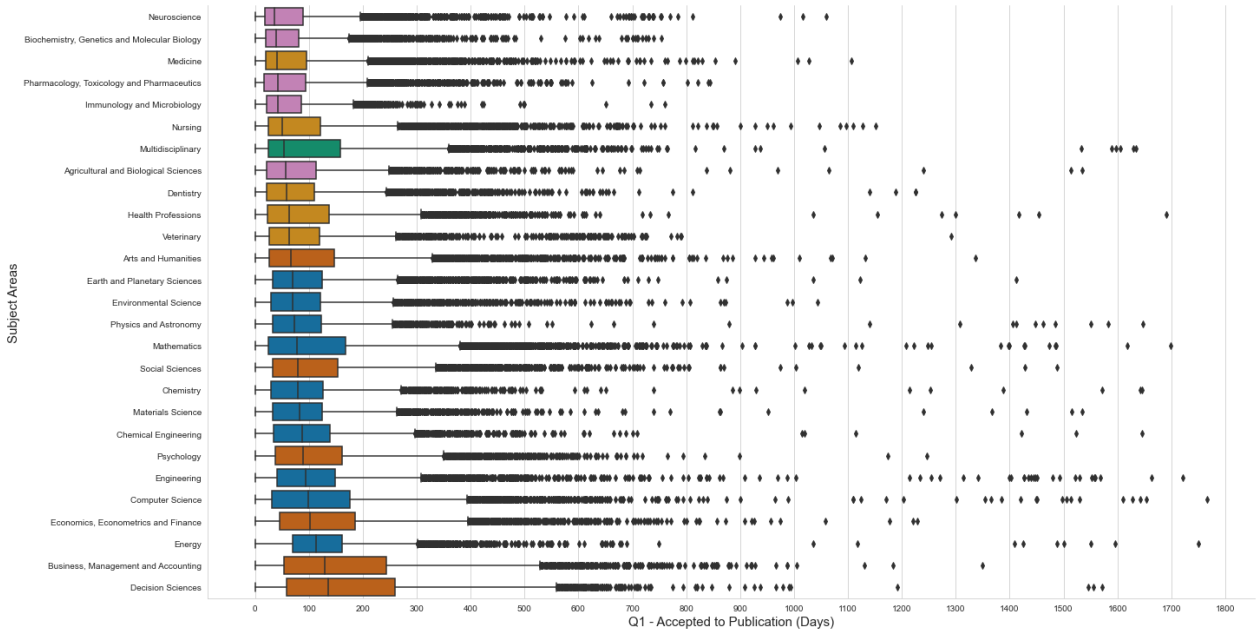


Figure 5: Total delay times in days across subject areas.

3.1.2. Published Year

Figure 6 displays the mean publication delay times across the years, displaying a steady (22%) decrease across the years. Pairwise one-sided t-tests for consecutive years indicate that the delay time improvements are only significant in half of the 10-year pairs, with only 2010 - 2011, 2011 - 2012, 2013 - 2014, 2014 - 2015 and 2015 - 2016 having significant difference ($p < 0.05$). This also shows that improvements in delay times were mostly in the first half of the 11-year period, 2016 being the last significantly improving year.

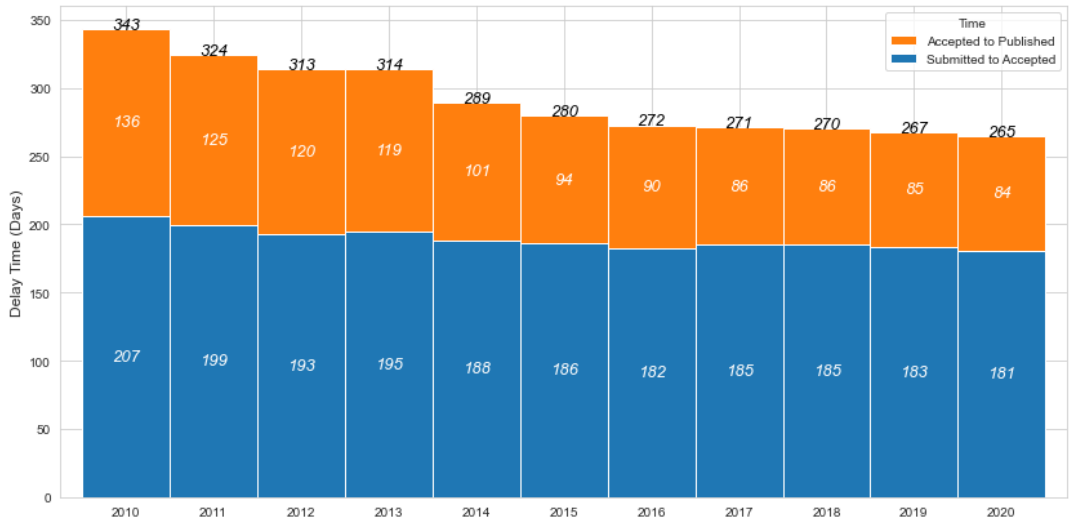


Figure 6: Average publication delay times breakdown in days across years. 2010 has the longest average delay with 343 days, and 2020 the shortest with 265.

Further analysis shows that this is mostly caused by the shorter publication delay times. This can be better viewed in Figure 7, displaying the yearly averages in line graph form. While acceptance times decreased from 136 to 84, showing a 38% improvement, at the same time span acceptance delays only improved by around 20 days, decreasing from an average of 207 to 181, a 12% improvement.

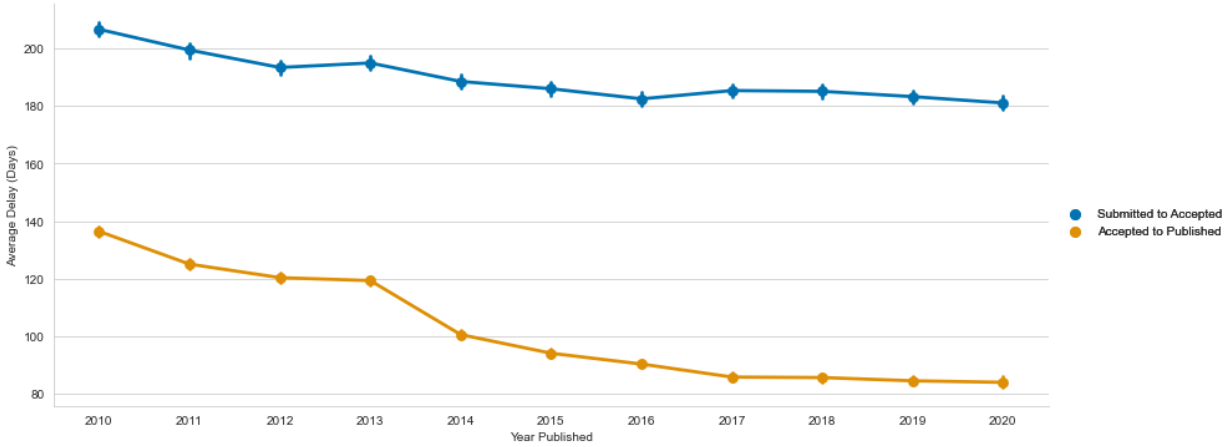


Figure 7: Average acceptance and publication delay times by publication year.

This is also supported by separate pairwise one-sided t-tests, as acceptance delay improvements are only significant for 4 consecutive year pairs (2010-2011, 2011-2012, 2013-2014 and 2015-2016) as opposed to publication delays with 6 (2010-2011, 2011-2012, 2013-2014, 2014-2015, 2015-2016 and 2016-2017).

3.1.3. Open Access

Figure 8 displays the number of articles released each year by non-OA and OA journals in the dataset. Throughout the 11-year period, OA article numbers show a steady increase with a total of 29,592 (14%) articles within the dataset. If we convert these numbers to journals, interestingly the same percentage of journals are OA with 3,775 (14%), whereas the remaining 23,559 (86%) journal records are either currently subscription based or were released before the journal went open access

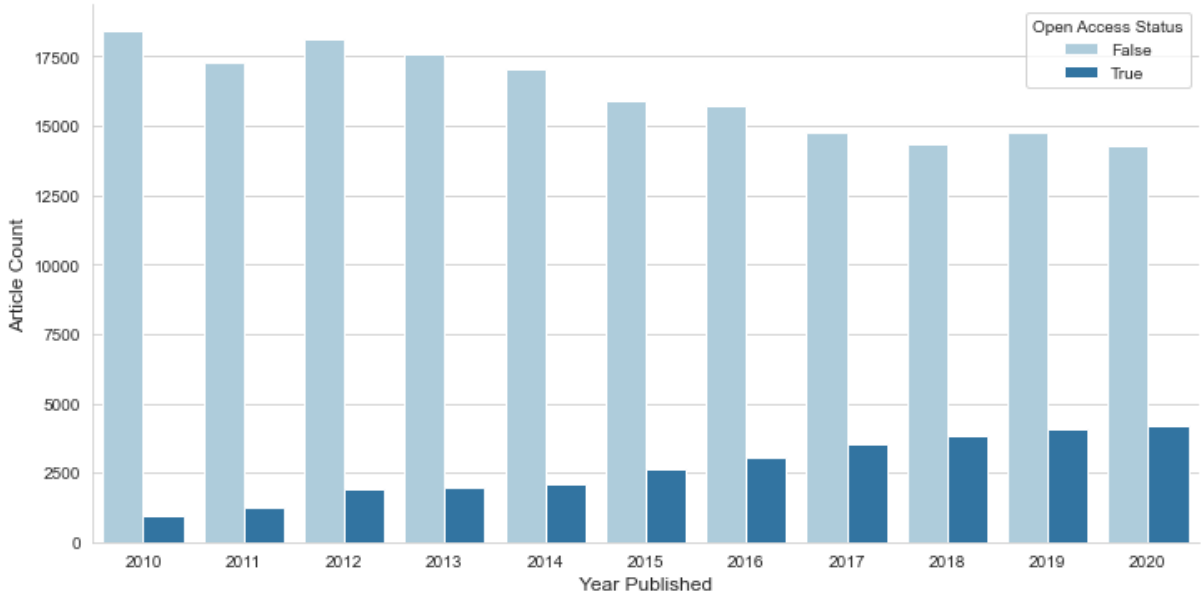


Figure 8: Number of OA and non-OA articles over the years, indicating the steady increase of Open Access publications over the years.

Publication delay times are also shorter for OA journals as they out-perform non-OA journals on both acceptance and publication delay times, as seen on Table 5. OA journals are especially fast in terms of publication delay with a median of just 27 days, or just under a month.

	Total Delay Time			Acceptance Delay			Publication Delay		
Open Access?	**Mean**	**Median**	**IQR**	**Mean**	**Median**	**IQR**	**Mean**	**Median**	**IQR**
No	310	253	163-392	200	153	89 – 256	110	81	33 – 143
Yes	190	152	97- 236	128	104	63 – 164	61	27	13 – 66

Table 5: Summary statistics for OA vs non-OA publication delay times

Figure 9 shows the yearly average delays for a more detailed view. OA journals not only have shorter delays times every year, but they also improved their delay performances, especially during 2010-2015 period, increasing the gap between.

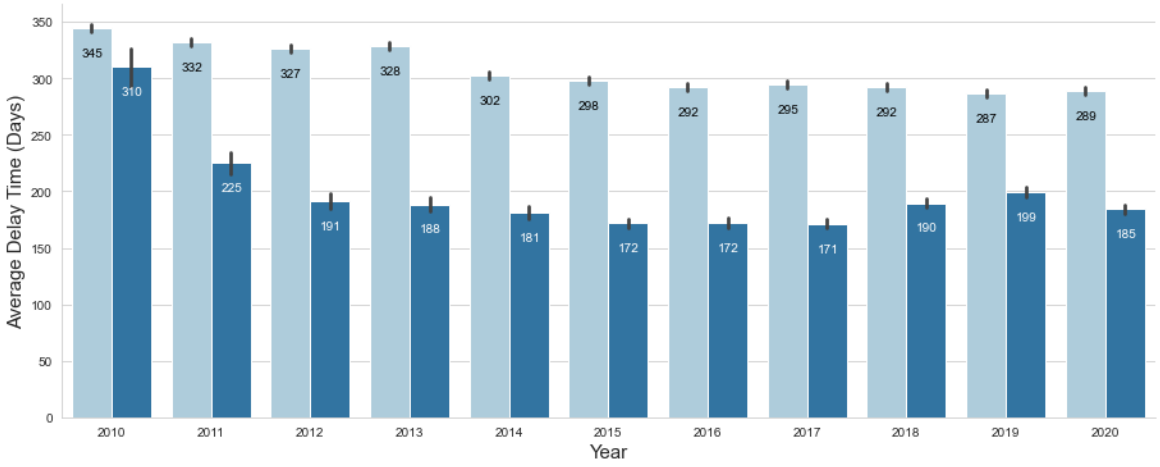


Figure 9: OA and non-OA total delay times by publication year.

The graph clearly shows the faster delay times of OA journals once again; however, it also shows that the improvements of OA journals start to stagnate starting with 2016 and delay times even start to increase with 2018 onwards.

3.1.4. Discussion of Descriptive Results

The results of the study show that the article delay times within the dataset are dependent on all three factors analyzed, and subject area, published year, and journal open access status play a significant role in determining their publication durations. Especially, main subject areas such as Life Sciences have a much quicker publication pipeline, whereas an average delay time of around 15 months seems to be the norm for Social Sciences. This in effect determines the speed of development of their domains, as well as the career progression of their authors. An Economics academic must work on and wait on average 3 times longer than a Chemistry academic to see the results of their hard work. Also from a different perspective, they must struggle three times more with submissions, revisions and waiting on publication pipelines to get their work out to their fellow academics.

These findings are all in accordance with the results of previous works mentioned. Björk et al. (Björk & Solomon, 2013) and Luwel (Luwel, 2020) reached the same conclusion in their analysis, using article data from or including the years in our analysis. This both supports and proves our claim in terms of subject area effect and shows that this is not year or publisher dependent pattern, but rather an accepted norm within the academic domain.

Publication year, although it also is a determining factor for publication delay times, can be considered a bit differently since it is not a factor neither authors nor publishers can directly control. The data indicates that the improvements in yearly average publication delays are mostly caused by the improvements in publication delay times. Although further research is necessary to completely come to a solution, most previous works on the topic mentions that this is due to the advancement of technology and the changes of publication methods, as digital publications allow articles to be shared very quickly, or even instantly for some cases, and without any page count limitations of a physical copy.

However, one very important and obvious improvement area is the acceptance times, as improvement on this delay is slower across the years and even increases between consecutive years in some cases.

It can be commented that any improvements for acceptance times needs a more structural and behavioral change from the publishers' side as technological advances and publication formats alone are not enough to drive a change for the better. As explained in Huisnan's work on the topic, with very little incentive on the reviewers' side and an overload of submissions, this naturally causes a backlog of articles to be reviewed and increases the acceptance delay times (Huisman & Smits, 2017).

The third factor addressed in the analysis; journal open access status also has an interesting effect. It can be clearly seen that the number of publications from OA journals are increasing at a rapid date each year, and the role of OA journals is very significant in the advancement of science as a whole and is very likely to continue to advance. They also have a significantly better performance when it comes to total delay times, which may be a supporting factor as to why authors prefer OA journals more and more. However, also in accordance with Powell's (Powell, 2016) findings**,** as the number of submissions to these journals are growing each year, they are also struggling to keep up with their performances.

3.2. Author Affiliated Country Analysis

3.2.1. Article Counts

The dataset contains articles from 184 affiliated countries in total. Figure 11 displays the number of articles from author affiliated continents, allowing a more compact view. It is seen that most of the articles in the dataset originate from three continents, Europe with 72,330 articles (making up for around one third of the entire dataset), followed by Asia and North America with 61,708 and 54,492 respectively.

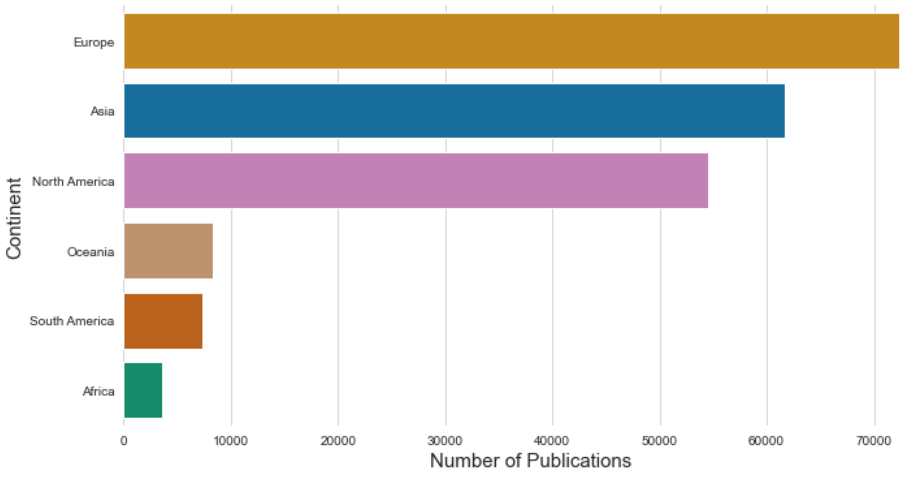


Figure 10: Total number of publications per affiliated continent.

Figure 12 displays the number of articles published originated from the top 25 affiliated countries. United States is the most common affiliated country among the authors, responsible for 22% of all and 84% of North American articles in the dataset. This is followed by China, United Kingdom, and Germany respectively, the only other affiliated countries with more than 10,000 articles. Africa is the only continent not represented by a country in this graph, as the country with the highest number of publications from Africa is South Africa with 895, ranking 35th.

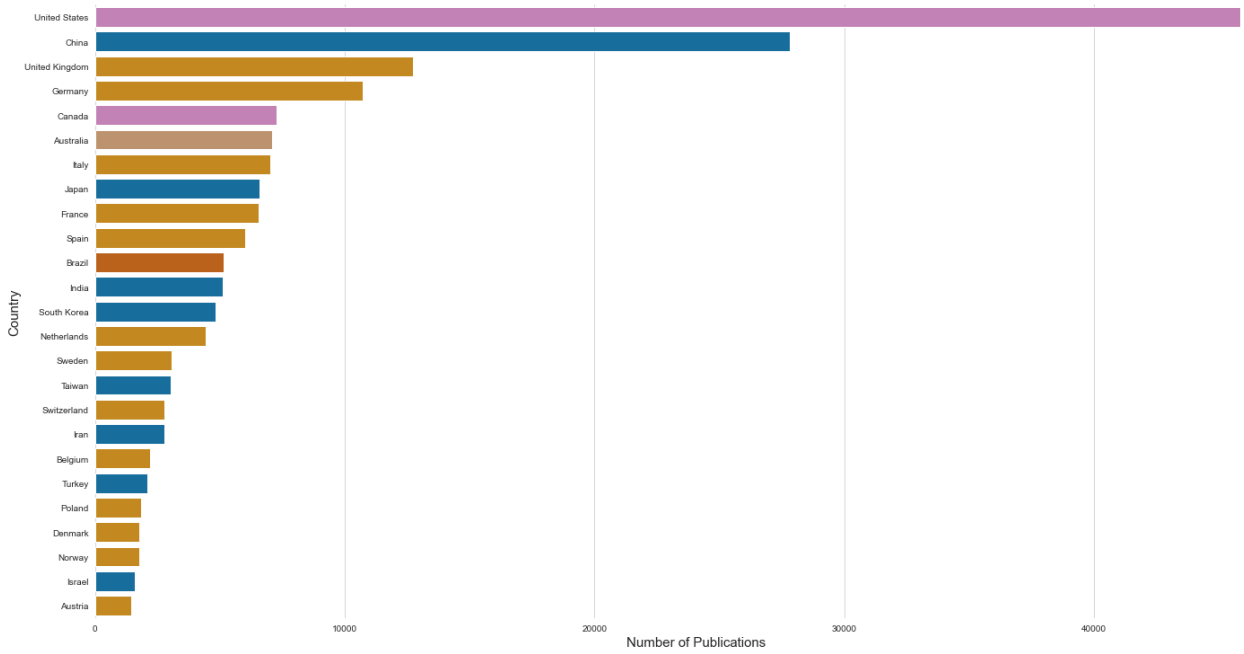


Figure 11: Total number of publications per affiliated country.

3.2.2. Publication Delay Times

Mean and median delay times for continents can be viewed in Table 6, indicating around a month's difference between the slowest and fastest continents. Asia has the lowest delay times overall, as it has the fastest time delays in terms of both mean and median. South America affiliated articles on the other hand seem to perform the worst overall, as they both have the largest median and second largest mean delay times. With only a 3 day difference North America is the slowest with 308 days on average, however, when the median delay times are also included it can be commented that "outlying" articles increase delay times are decrease the overall performance, especially for North America and Europe.

Continent	Median Total Delay Time	Mean Total Delay Time
Africa	247	291
Asia	**226**	**276**
Europe	238	294
North America	245	**308**
Oceania	244	292
South America	**254**	305

Table 6: Mean and median total delay times by continents. Asia is the fastest overall, while North and South America affiliated articles suffer from longer delays.

Figure 13 presents the total delay time plots from the top 25 most affiliated countries. Japan has the overall shortest delay times and is the only country with a median value lower than 200 days. China follows this with a median of 216 days, making them the fastest published among the top 5 most affiliated countries. On the other side of the spectrum, there are France, Brazil, and Turkey as the

overall slowest countries. Comparing the best and worst performers in this table shows that there is a difference of 78 days on average between the best performing Japan and worst performing Turkey.

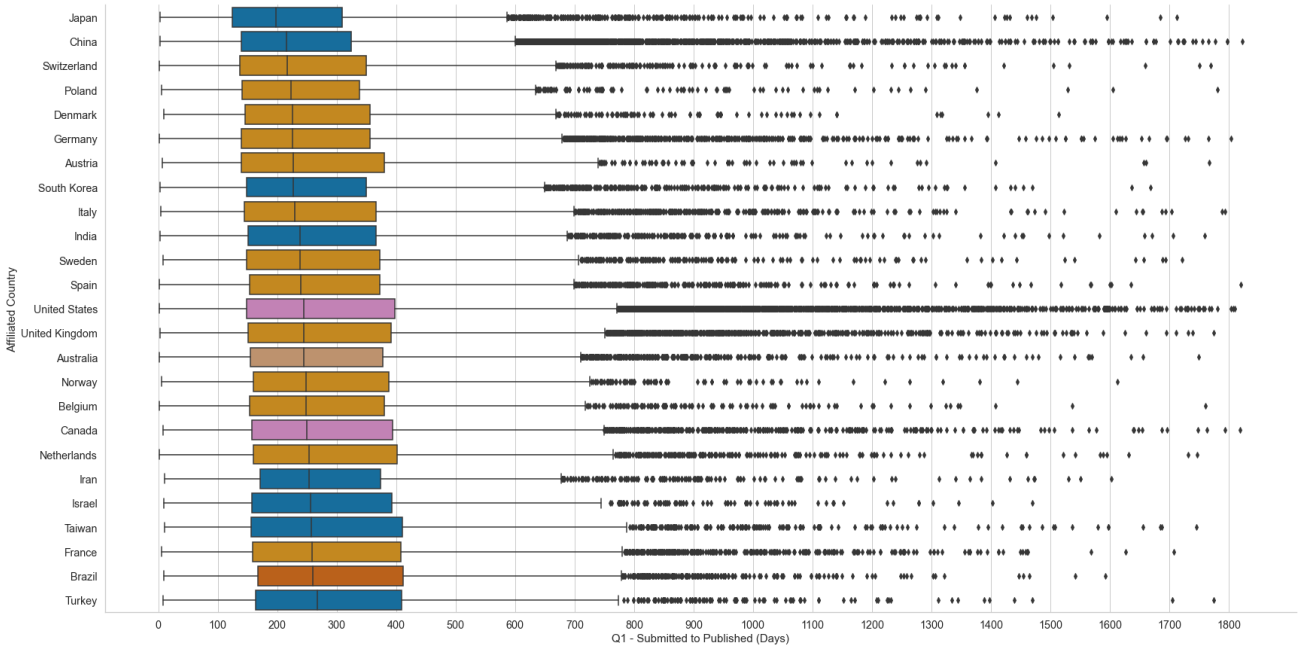


Figure 12: Top 25 most affiliated countries total delay times boxplot. Japan and China, both from Asia, are have the fastest median times overall. The rest of the graph is dominated by mostly European countries.

3.2.3. Global North vs. Global South

In the dataset, articles affiliated with a global north country make up 74%, equaling to 153,145 articles in total. In comparison, the number of global south originated articles are 54,640. There are 58 different countries represented within the Global North group with United States, United Kingdom, Germany, and Canada being the most represented, while there are 126 countries in the global south group, with China, Brazil, India, and Iran as the top runners. It should be noted that there are several countries originated from Global South, such as Angola and Virgin Islands that have a single article associated with them, increasing the number of total countries in the group.

Table 7 displays the summary statistics for the matched dataset. From a first glance it can be commented that both mean and median values are quite similar, with only a couple of days difference between the two groups. The average delay time for both groups is around 280 days, with around 180 of those days (or 6 months) spent on acceptance delays, while the rest 100 delay days are due to publication.

	Total Delay Time		Acceptance Delay		Publication Delay	
	Mean	**Median**	**Mean**	**Median**	**Mean**	**Median**
Global North	281	228	179	135	102	73
Global South	279	230	176	134	103	77
Difference (days)	**2**	**-2**	**3**	**1**	**-1**	**-4**

Table 7: Summary statistics for Global North vs. Global South matched dataset. The delay times indicate a similar publishing experience for both groups.

Moving on with the detailed analysis and following the same order as before, Figure 13 displays the total delay times with respect to the main subject areas. Here, the delay times are quite similar for three of the five main subject areas, with Life Sciences being the highest among them with a 5 days' difference. However, Multidisciplinary journals seem to have shorter delay times for articles originated from Global North, as the average delay times are 25 days shorter. In contrast, Social Sciences articles tend to have a shorter delay time if they are affiliated with a Global South country, as Global North affiliated articles have 25 days longer delay times on average.

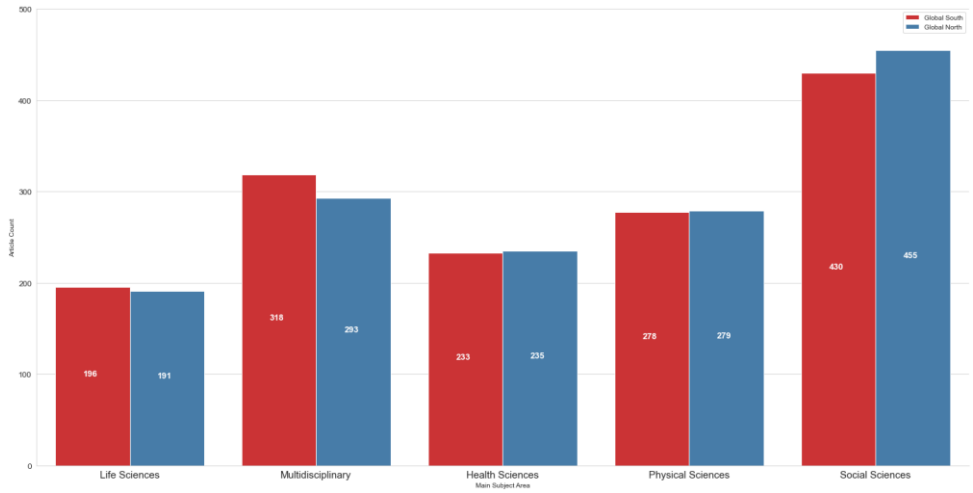


Figure 13: Global North vs. Global South average total delay times with respect to main subject areas. Multidisciplinary journals seem to publish Global North articles faster, where Social Sciences journals are faster for articles affiliated with Global South.

Figure 14 gives a bit more context to our findings from above, as it displays the subject area delay times for both groups. We can observe that the difference in Life Sciences delay times is mostly originated from *Arts and Humanities* and *Psychology* articles.

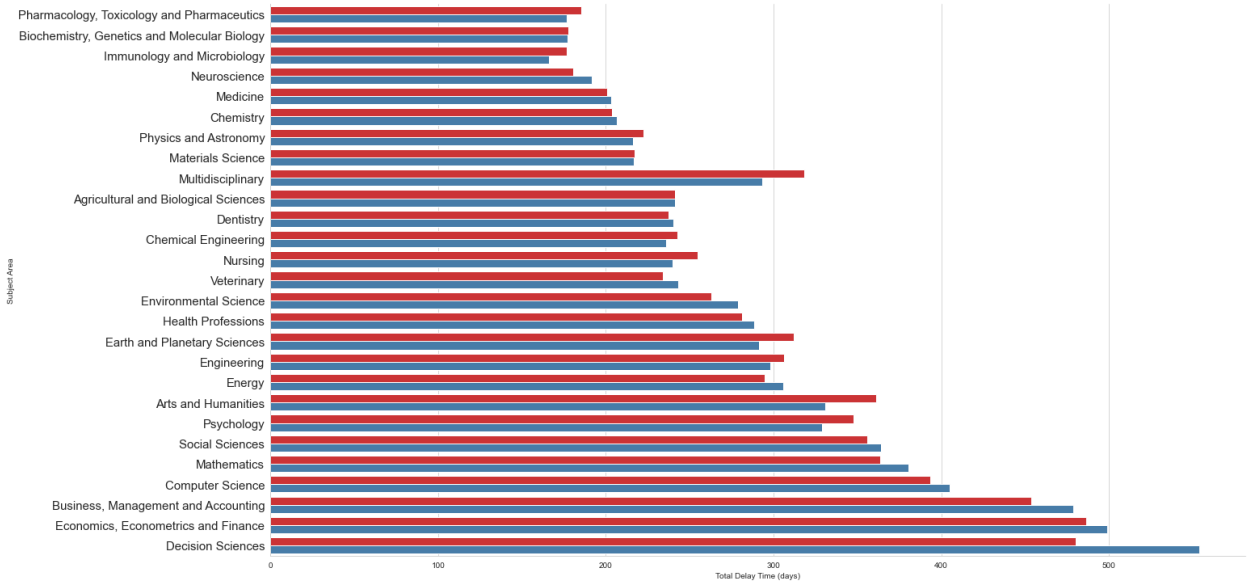


Figure 14: Global North vs. Global South average total delay times with respect to subject areas.

The yearly averages in Figure 15, indicate a minor but interesting pattern. Although the difference is quite small, Global North articles seem to have shorter delay times between 2010-2014. For the following 6 years the roles are reversed, and Global South articles consistently have shorter delay times. We can also observe that delay times have decreased for both groups over time, in accordance with our analysis earlier.

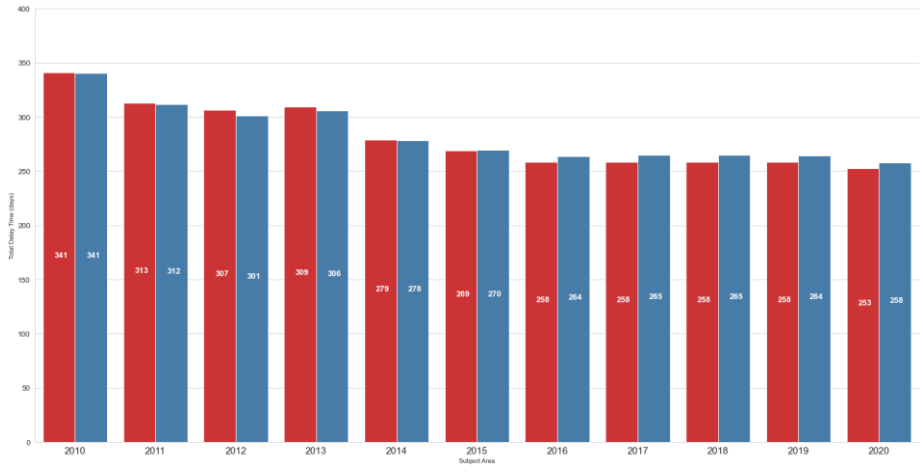


Figure 15: Global North vs. Global South yearly average delay times. Global North articles have shorter delays until 2015, where the roles are reversed.

Figure 16 compares the journal Open Access status performances between the groups. OA journals have faster delay times overall, but they are also around 2 weeks shorter for Global North associated articles, with an average of 214 days compared to Global South's 229. However, non-OA journals instead have shorter delays for Global South affiliation, as Global North affiliated articles have around 1 week longer delay times in comparison.

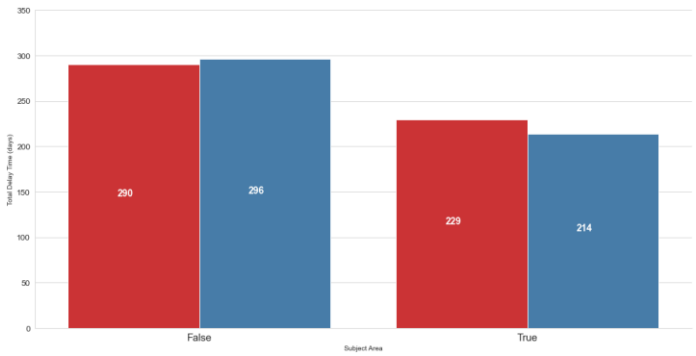


Figure 16: Global North vs. Global South journal Open Access status average delay times. OA journals are faster for Global North affiliated publications, but non-OA journals are faster for Global South.

Finally, we can use the "Publisher Origin" to compare the two groups, as seen on Figure 17 below. Publisher originating from United Kingdom, Europe, Asia, South America, Africa, and Oceania have lower delay times, although Asia, South America and Africa are all part of Global South (as Australia is considered a part of Global North). United States, Netherlands, Germany, and North America based publishers have shorter delay times for Global South affiliated articles, although all these countries/regions are part of the Global North.

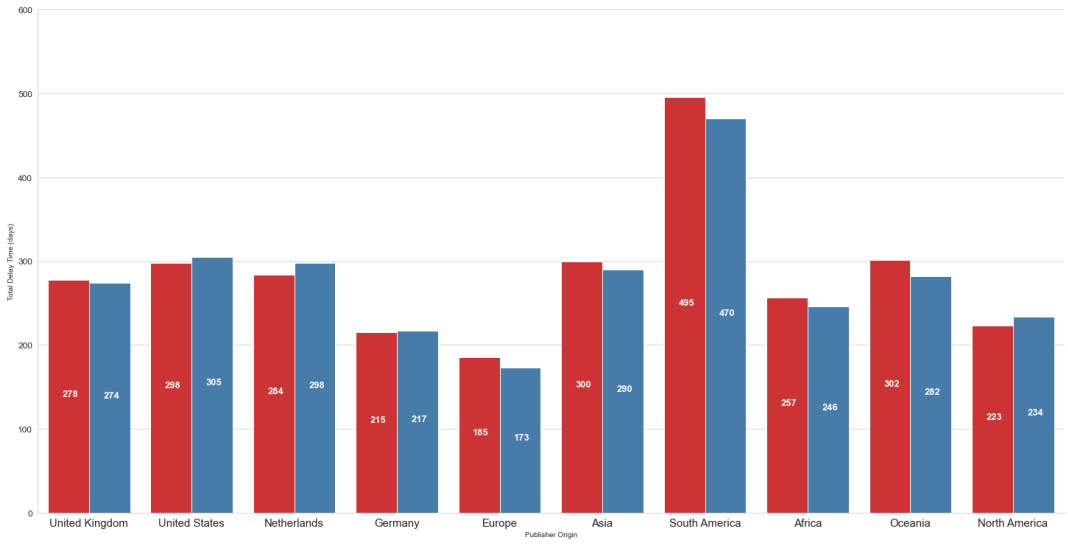


Figure 17: Global North vs. Global North, average publisher origin delay times. South American publishers seem to have the biggest delay gap between the two groups with 25 days, while Germany is the smallest with an average difference of 2 days.

3.2.4. Native English vs. Non-native speakers

The total dataset contains 13 affiliated countries that have English as their first language. However, as these countries include United States, United Kingdom, and Canada, they make up the 36% of the entire dataset.

Table 8 shows the summary mean and median values for the two matched groups, clearly indicating a similarity across the delay times. All values are very similar to each other with a maximum difference of just 2 days.

	Total Delay Time		Acceptance Delay		Publication Delay	
English as first language?	**Mean**	**Median**	**Mean**	**Median**	**Mean**	**Median**
True	306	245	200	151	106	71
False	304	245	201	153	104	70
Difference	**2**	**0**	**-1**	**-2**	**2**	**1**

Table 8: Summary statistics for native and non-native English-speaking affiliated countries. The graph shows the delay time statistics for both groups are very similar to each other.

Figure 18 shows the average total delay times for “native” and “non-native” speaking countries from main subject areas perspective. The graph showcases the similar delay performances, however native English affiliated articles seem to have shorter delays for Multidisciplinary journals, whereas Health Sciences journals are faster for non-native affiliated articles with 7 days, or exactly 1 week.

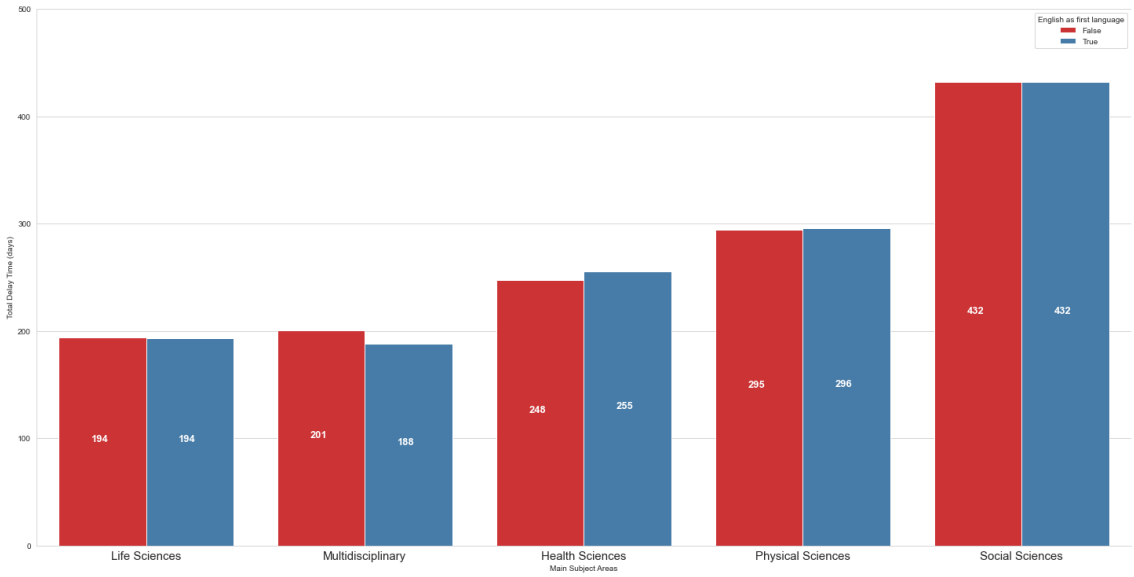


Figure 18: Native vs. non-native Englisher speaking countries average total delay times with respect to main subject areas. The delays are quite similar for the two groups, with only minor differences in Multidisciplinary and Health Sciences publications

Another interesting observation can be made for the journal origin section in Figure 19, as it shows clear distinctions between the two groups for certain publication regions. Here, delay time differences are minor for most publishers and vary around a week, but South American and Oceania based publishers both have significantly shorter delay times for native English affiliated articles. The delay times are shorter by 67 days and 39 days respectively for native speakers in these regions. This is also intriguing as countries from South America are considered as “non-native” English speakers, whereas Oceania countries are generally considered as native English speaker.

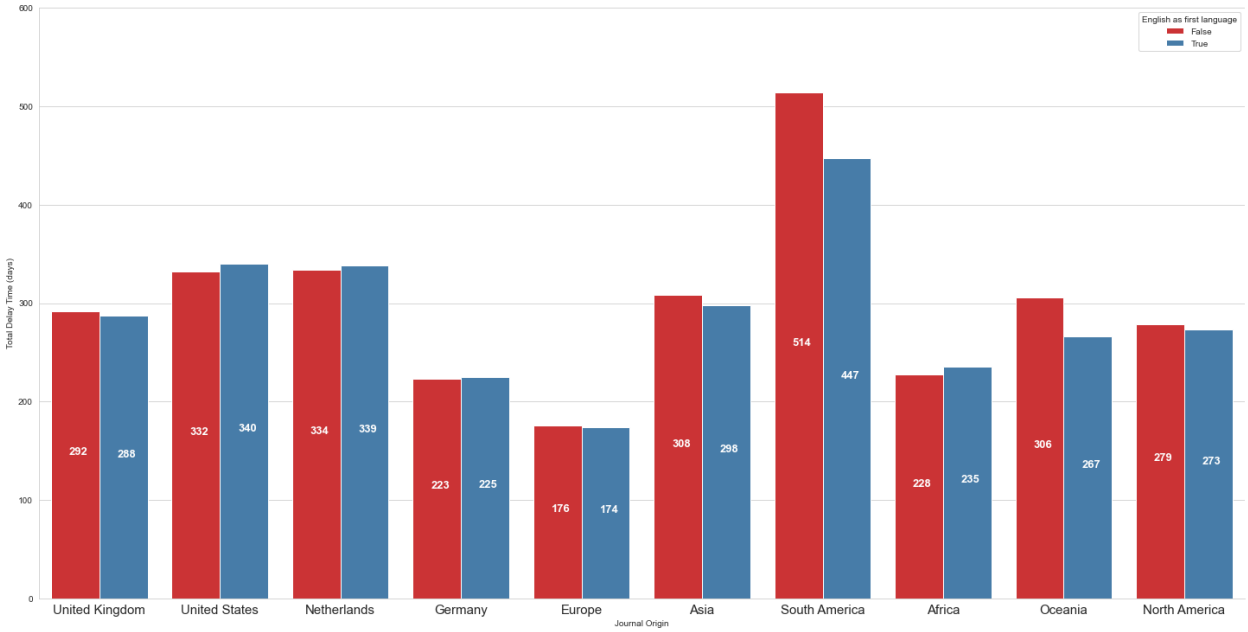


Figure 19: Native vs. non-native Englisher speaking countries average publisher origin delay times. South American publishers seem to have the biggest delay gap between the two groups with 25 days, while Germany is the smallest with an average difference of 2 days.

3.2.5. G7 Countries vs. Rest of the world

The total dataset has 31 G7 member affiliated countries. As all the member countries are also part of Global North group discussed earlier, this is a more exclusive group, but they still make up most of the article dataset with 60%.

Table 9 below shows the mean and median publication delay times for G7 and non-G7 affiliated countries matched dataset. Following the results above, the two groups display quite similar performances, as they have an average total delay difference of just three days, which seems to originate from the acceptance delay times.

	Total Delay Time		Acceptance Delay		Publication Delay	
G7 Member?	**Mean**	**Median**	**Mean**	**Median**	**Mean**	**Median**
True	287	233	184	139	103	73
False	284	232	181	138	103	76
Difference	**3**	**1**	**3**	**1**	**0**	**-3**

Table 9: Summary statistics for G7 member and non-member affiliated countries. The average delay times are again quite similar for both groups, with a 3-day total difference.

The similarity also continues for main subject areas, with the only delay time differences originating from Multidisciplinary and Social Sciences submissions. Non-G7 member affiliated countries seem to have faster publication processes for Multidisciplinary articles, while they suffer 16 days more on average for their Social Sciences submissions.

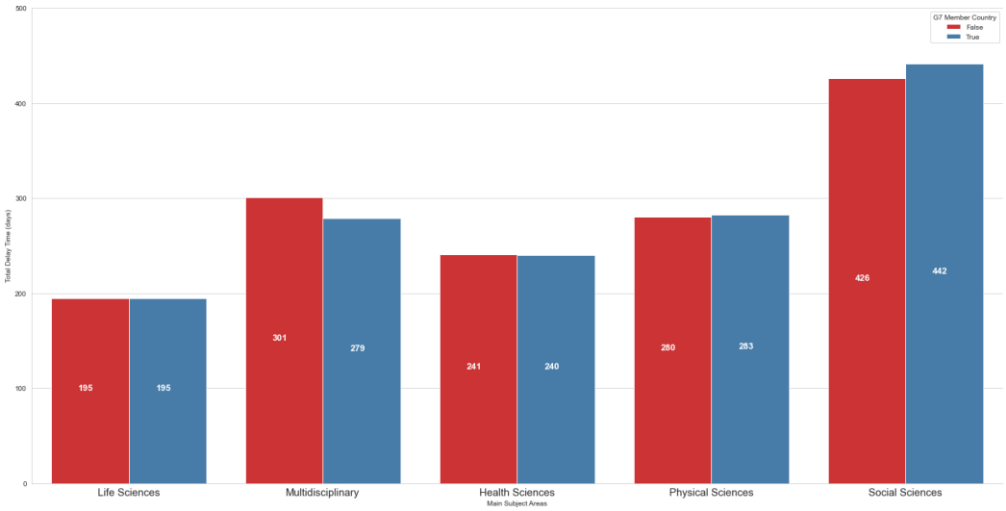


Figure 20: G7 member vs. non-member affiliated countries average total delay times with respect to main subject areas.

Figure 21 shows the comparison results from the other analysis perspectives, all indicating a performance similarity between the groups. The findings are also in accordance with the Global North. Vs South perspective earlier, as they both follow similar patterns albeit some minor differences.

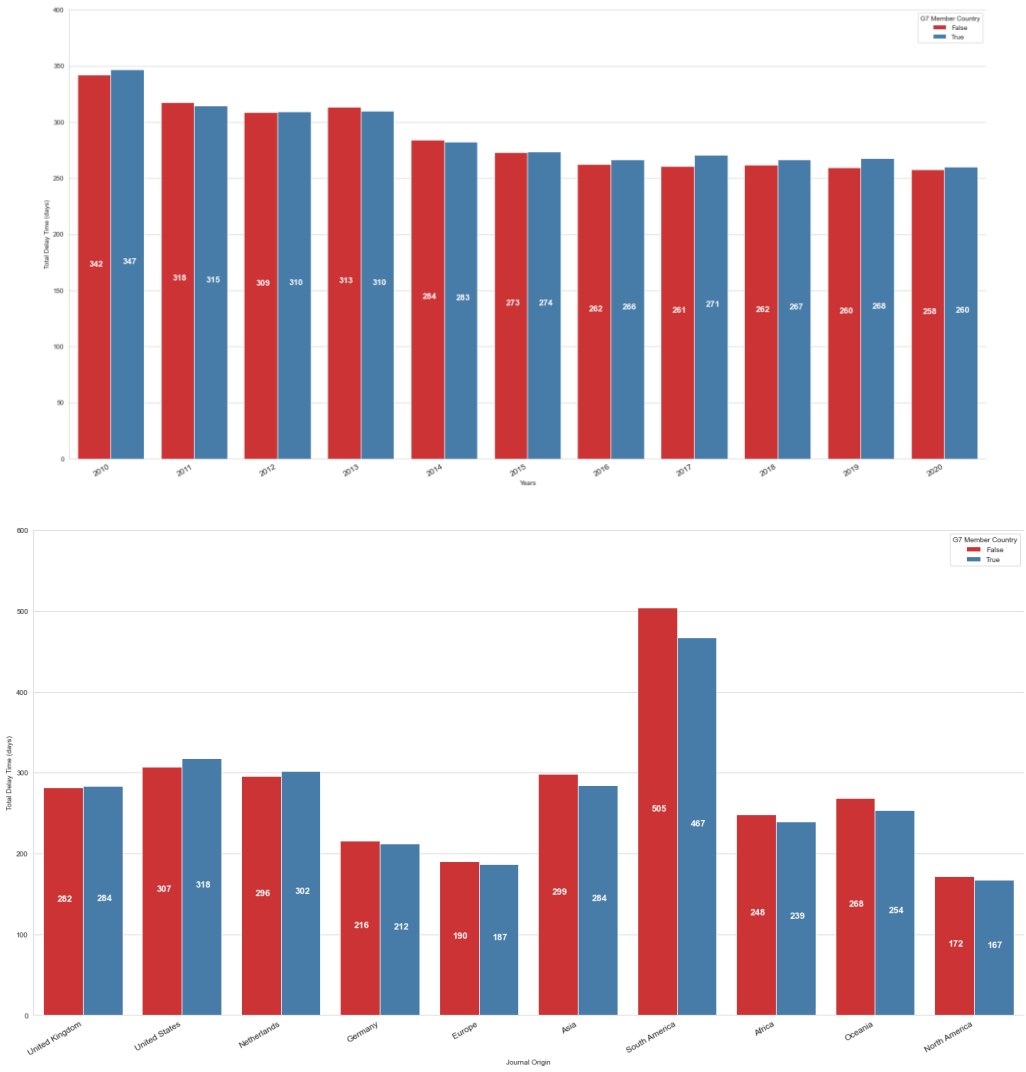


Figure 21: G7 member vs. non-member yearly and publisher origin average delay graphs. Both graphs follow a similar pattern to the Global North vs. Global South analysis earlier.

3.2.6. Discussion of Country Affiliation Analysis

The results of the analysis show that there is great disparity of representation between certain countries and continents. Although the data collection did not have any limitations on an article's country affiliation, Europe, Asia, and North America affiliated articles make up around 90% of the entire dataset. On the other end of the spectrum the remaining three continents, Oceania, South America, and Africa, can only account for the remaining 10%.

Asia not only has the second most affiliated articles but also has the fastest delay times overall, where Japan, China and Russia significantly improve the continent's overall performance. South American affiliation on the other hand, affects an article's publishing process negatively, as these articles suffered around 30 days more than Asia affiliated articles and 13 days more than the overall average of 292 days. It is also interesting that United States affiliation, which is the most common country in the entire dataset, does not offer a better publication experience to its authors.

The three matching analyses performed present a different situation. None of the analysis approaches resulted in a significant difference between the groups, with only a 2-3 day difference in between. Especially affiliation with English as the first language seem to not affect a paper's publication process. This is understandable as the journals included in this study are the most prestigious at their domain, and submissions made to these journals can be expected to have high quality both in term of study performed and their stylistic and linguistic capabilities.

Global North vs. South approach can be analyzed together with the G7 member analysis, as the G7 member countries are a more exclusive group of Global North countries. Main subject area, year and Open Access comparisons all show that despite minor differences, the publication experience is the same for the differently affiliated authors. One note is that Open Access journals seem to favor the "slower" group in their publications. This is also fitting for their "alternative solution" approach to the publication domain.

For both groupings, publication origin has the biggest differences, as South America a G7 non-member and part of Global South, has significantly shorter delays for the opposite group. The rest of the publisher origins seem to have a fair distribution between the groups, except for Oceania publishers which seem to publish Global North / native-English and G7 member affiliated articles faster in general.

4. CONCLUSION

The main objective of this study was to answer the question of "How do publication speeds differ?" from different perspectives to define and analyze the academic publication landscape. However, I wanted this study to differ from other papers on the issue. As the publication delay times are not a readily available metadata field the previous works on the topic are based on a single publisher or database to collect their data. In this study, I followed a different approach and the most crucial data field the study relies on was collected by manually checking and retrieving submission, acceptance, and publication times from the articles with a tedious journal scanning and date gathering pipeline. By scanning all available resources and by combining the data gathered, I believe I accomplished the purpose of creating an inclusive (in terms of publisher, article, and author details) and representative dataset. The data collection pipeline and the resulting dataset is not only useful for study, but it also has the potential to be a baseline for any future works on the topic, simply by including new analysis perspectives (such as PlumX metrics) or increasing the scope (in terms of years, author details such as gender or including Q2 to Q4 quartile journals). The sampling design is flexible and allows new data fields to be added.

Article delay times in the dataset are also not a single time value, but they include acceptance and publication delays separately. This allowed me to analyze the steps of publication in more detail. It is seen that the acceptance delays are the main cause of the issue, as both previous studies and this study indicate that acceptance times cause most of the total delay time and are not improving at the same rate as publication delays. With many previous works on the topic highlighting the significance of peer-reviewing, I also believe that a gate-keeping policy is required to maintain the quality of submissions. However, without significant improvements on the process itself, the data shows that technological advances alone are not enough to continuously improve publication delay experience. The instant availability seems to quicken the time between acceptance and publication but in terms of acceptance delays and peer-reviewing, both publishers and authors are stuck with frustratingly long waiting times. This is further supported as the 11 years included in our analysis show little improvement on this aspect within the time frame.

The data collected also aimed to answer the research question from the perspective of Subject Area, Publication Year, journal Open Access status and first author affiliated country. The results show that there is a clear distinction of delay times between different subject areas. This is completely in line with the previous works on the topic mentioned. Life and Health Sciences articles have significantly faster publication processes, while Social Sciences suffer the most. This alone is an interesting conclusion, however including 2021 articles in the dataset can also be a great addition to this approach. Effects of the COVID-19 pandemic on the publication speeds of these subject areas can give us more insight on the underlying reasons and showcase the performances in case of extreme circumstances in a scientific domain. However, in general these delay times are likely to be accepted as the norms of their respective domains, and likely to be affected directly by the number of submissions, active academics, and new reported findings on the domain. This can be further analyzed and supported by including the rejected submissions to the publishers.

Open Access journals seem to offer the alternative they claim to be for academic publication, at least for the current time being. Their overall faster delay times have increased their popularity over time and increased their publication ratios steadily. However, they are still a minority compared to the

traditional publishing giants, and they also displayed performance stagnation and drops along with their increasing popularity. It seems that they are offering a faster and easier experience overall, but it is also important to follow their experience in the following years to see if they can maintain their performance. If they can maintain lower delays with more publications, they can be the enabler of change in the publication domain and cause traditional publishers to rethink about their publication pipelines.

Author country affiliation analysis confirms a well-known and intuitive idea regarding the academic landscape. In a completely random data collection pipeline including only the most prestigious (Q1 ranked in SJR) journals, both the publishers and author spaces are dominated by several countries. This is understandable as these countries offer more professional opportunities to academics and have long standing and prestigious institutions. However, this domination also means that both what is being published, and the scope and subject of the papers themselves are directly affected by these counties' decisions. They in effect direct and determine the entire academic domain. This also means that from an author's perspective moving or working in one of these countries offers an easier and quicker way to achieve their goals.

The matching analysis shows little to no difference between the analysis groups, indicating that there is no unfair gatekeeping taking place by the publishers due to an author's affiliation. This is good news for the entire academic domain and for any future publications. A submitted paper should be considered regardless of the affiliated counties' geographical, cultural, or economic situation and paper quality alone should be the deciding factor. This analysis shows that at the time being and for the analyzed publications, these factors are not significantly determining. This can also be explained by the stature of the journals and the authors making the submissions. As this dataset includes the highest ranked journals, they are also expected to receive submissions from the most prestigious academics of their domains, and in this high-quality environment, scientific prowess is more deterministic than potential personal biases.

5. LIMITATIONS

The limitations of this project can be approached from each of the journal, article and author datasets used during analysis. From a journal point of view, the biggest limitation of the project is the journal SJR quartile rankings. As this study only includes Q1 journals, it only contains the highest ranked journals from each scientific domain. By adding more quartiles, we would be able to not only increase the scope of the project to get a more general picture, but also would be able to analyze the delay times with respect to journal rankings.

Article dataset also causes several setbacks limiting the scope of the analysis. As the project uses CrossRef platform to collect most of the article metadata, the “references” and “is referenced by” columns are not regularly updated and were not included during the analysis. A different or additional database to collect this information would enable a deep dive approach and would also offer more insight during the matching analysis. Several previous works on the topic mention an inverse correlation between an article’s delay times and its subsequent received references. This would make the matching to be more precise in terms of article impact and would allow us to compare all analysis perspectives from a different angle.

The study also only includes first authors from each article due to the size of the total author dataset and the API limitations while gathering affiliation information. This still allowed us to conduct an affiliated country analysis, however by including all the authors of a paper, we would be able to detect any underlying patterns. As authors from different affiliations can work together on a paper, a specific author’s affiliation may have a significant role in determining its publication fate. For example, comparing South American affiliated first authors based on the remaining author affiliations can give more insight on the overall slow performance of the continent.

6. REFERENCES

Asaad, M., Rajesh, A., Banuelos, J., Vyas, K., & Tran, N. (2020). Time from submission to publication in plastic surgery journals: The story of accepted manuscripts. *Journal of Plastic, Reconstructive & Aesthetic Surgery, 73*, 383-390.

Björk, B.-C. (2019). Acceptance rates of scholarly peer-reviewed journals: A literature survey. *El profesional de la información, v. 28, n. 4, e280407*, 1-9.

Björk, B.-C., & Solomon, D. (2013). The publishing delay in scholarly peer-reviewed journals. *Journal of Informetrics*, 915-923.

Clapham, P. (2005). Publish or Perish. *BioScience, Volume 55, Issue 5,* , 390–391.

De Rond, M., & Miller, A. (2005). Publish or Perish Bane or Boon of Academic Life? *JOURNAL OF MANAGEMENT INQUIRY, Vol. 14 No. 4*, 321-329.

Gerring, J. (2012). Mere Description. *British Journal of Political Science, 42(4*, 721-746.

Hendricks, G., Tkaczyk, D., Lin, J., & Feeney, P. (2020). Crossref: The sustainable source of community-owned scholarly metadata. *Quantitative Science Studies, 1(1)*, 414–427.

Herbert, R. (2019). Accept me, accept me not: What do journal acceptance rates. *ICSR PERSPECTIVES*, 2-11.

Hill, R., & Stein, C. (2021). Race to the Bottom: Competition and Quality in Science. *Working Paper*, 1-77.

Himmelstein, D., & Powell, K. (2016). Analysis for "the history of publishing delays". *blog post v1.0 (v1.0). Zenodo*.

Hsiehchen, D., & Espinoza, M. (2016). Detecting editorial bias in medical publishing. *Scientometrics, 106*, 453–456.

Huisman, J., & Smits, J. (2017). Duration and quality of the peer review process: the author's perspective. *Scientometrics, 113*, 633–650.

Larivière, V., Haustein, S., & Mongeon, P. (2015). The Oligopoly of Academic Publishers in the Digital Era. *PLoS ONE 10(6): e0127502.* , 1-15.

Lee, Y., Kim, K., & Lee, Y. (2017). Publication Delay of Korean Medical Journals. *The Korean Academy of Medical Sciences, 32*, 1235-1242.

Luwel, M. v. (2020). Characteristics of Publication Delays Over the Period 2000–2016. *Evaluative Informetrics: The Art of Metrics-Based Research Assessment*.

Lyman, R. L. (2013). A Three-Decade History of the Duration of Peer Review. *Journal of Scholarly Publishing*, 211-220.

Mrowinski, M., Fronczak, A., & Fronczak, P. (2020). The hurdles of academic publishing from the perspective of journal editors: a case study. *Scientometrics, 125*, 115–133.

Neill, U. S. (2008). Publish or perish, but at what cost? *The Journal of Clinical Investigation, Volume 118, Number 7*, 2368.

Nicholas, D., WATKINSON, A., JAMALI, H., HERMAN, E., TENOPIR, C., VOLENTINE, R., . . . LEVINE, K. (2015). Peer review: still king in the digital age. *Learned Publishing, 28*, 15–21.

Nishikawa-Pacher, A. (2022). Who are the 100 largest scientificpublishers by journal count? A webscraping approach. *Journal of Documentation Vol. 78 No. 7*, 450-463.

Powell, K. (2016). The Waiting Game. *Nature, Vol 530*, 149-151.

Reif, F. (1961). The Competitive World of the Pure Scientist: The quest for prestige can cause conflict between the goals of science and the goals of the scientist. *Science Vol 134, Issue 3494*, 1957-1962.

Siler, K., Lee, K., & Bero, L. (2015). Measuring the effectiveness of scientific gatekeeping. *PNAS, vol. 112, no. 2*, 360-365.

Strevens, M. (2003). The Role of the Priority Rule in Science. *The Journal of Philosophy, Volume 100, Issue 2*, 55-79.

Taskin, Z., Taskin, A., Dogan, G., & Kulczycki , E. (2022). Factors affecting time to publication in information science. *Scientometrics*, 1-17.

van Lent, M., Overbeke, J., & Jan Out, H. (2014). Role of Editorial and Peer Review Processes in Role of Editorial and Peer Review Processes in Eight Medical Journals. *PLoS ONE 9(8): e104846.*

Zhu, H. (2021). Home country bias in academic publishing: A case study of the New England Journal of Medicinev. *Learned Publishing*, 1-7.

APPENDIX

	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Agricultural and Biological Sciences	47465	52119	57737	64053	65655	67758	66054	68190	72565	73689	83335
Arts and Humanities	18460	19887	22910	24953	27587	26795	27473	28467	31940	36163	42901
Biochemistry, Genetics and Molecular Biology	68496	72196	77766	80790	80560	77869	74461	75934	73678	75108	75142
Business, Management and Accounting	9418	10401	11052	10901	10900	11473	12854	13108	14365	17127	20896
Chemical Engineering	15673	17279	19260	22295	25703	29249	30773	30860	32615	34737	37545
Chemistry	45387	51188	52839	54117	60266	63302	66037	64596	65267	72178	79377
Computer Science	18820	23264	25476	23748	26127	27769	29552	32932	37725	45886	57098
Decision Sciences	2239	2376	2963	3007	3123	3080	3662	3664	4003	4408	5915
Dentistry	4121	4885	5107	4995	5264	5232	4964	5510	5772	5886	6493
Earth and Planetary Sciences	28876	30142	32606	34190	36707	36969	38718	39955	42574	44313	49786
Economics, Econometrics and Finance	6043	6387	7457	7237	7288	7582	7513	8262	9309	10552	13794
Energy	11745	13867	14063	16648	19550	18986	21572	22078	23759	28387	32921
Engineering	59039	64576	68101	73500	78505	82943	91812	96380	113405	122349	125975
Environmental Science	24562	26503	28695	31612	34052	34950	36172	36449	46308	48053	55698
Health Professions	7090	7514	7812	8367	8586	8778	8678	9127	8985	9906	11074
Immunology and Microbiology	13142	14064	15040	14778	14505	13726	13999	14677	15395	15659	17393
Materials Science	45257	52565	53193	55377	59524	61487	63768	63791	71619	74176	82416
Mathematics	20244	24328	24552	24570	23240	24805	27311	26143	27708	30625	34222
Medicine	248490	260927	273565	289150	291525	299405	307435	310306	312211	304500	352515

Multidisciplinary	11026	11088	11752	13742	16783	21518	31853	36019	32295	50318	57500
Neuroscience	14553	15471	16340	16238	16119	16960	17679	18309	18067	17893	17402
Nursing	10746	10294	11250	11550	11881	12182	12282	13450	13900	14812	18421
Pharmacology, Toxicology and Pharmaceutics	20655	24182	25770	23276	24264	23546	22894	23369	26892	31441	32073
Physics and Astronomy	60592	63800	65702	67266	68003	62534	63217	59890	69345	75807	72274
Psychology	9848	10970	11875	13951	15207	15054	15883	16897	17979	19692	18686
Social Sciences	34414	37837	40405	44437	48156	48022	50479	55951	60921	70720	102759
Veterinary	5322	5499	6141	6023	5553	5806	5576	5707	6843	7201	11079

Appendix 1: Q1 journals - Total document count matrix

	2010	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**
Agricultural and Biological Sciences	593	593	594	595	595	595	595	595	595	595	596
Arts and Humanities	581	583	585	586	587	587	587	588	589	590	592
Biochemistry, Genetics and Molecular Biology	595	595	596	596	596	596	595	596	595	595	596
Business, Management and Accounting	564	568	569	569	569	570	574	574	576	580	584
Chemical Engineering	578	580	582	585	587	588	589	589	589	590	591
Chemistry	592	593	594	594	594	595	595	595	595	595	596
Computer Science	582	585	586	585	587	588	588	590	591	593	594
Decision Sciences	474	479	499	501	504	502	516	516	522	528	545
Dentistry	524	535	537	536	539	539	536	541	544	545	550
Earth and Planetary Sciences	588	589	589	590	591	591	591	591	592	592	593
Economics, Econometrics and Finance	546	549	556	554	555	556	556	560	564	568	575
Energy	571	575	576	579	582	582	584	584	585	588	590
Engineering	594	595	595	595	596	596	596	597	597	597	597

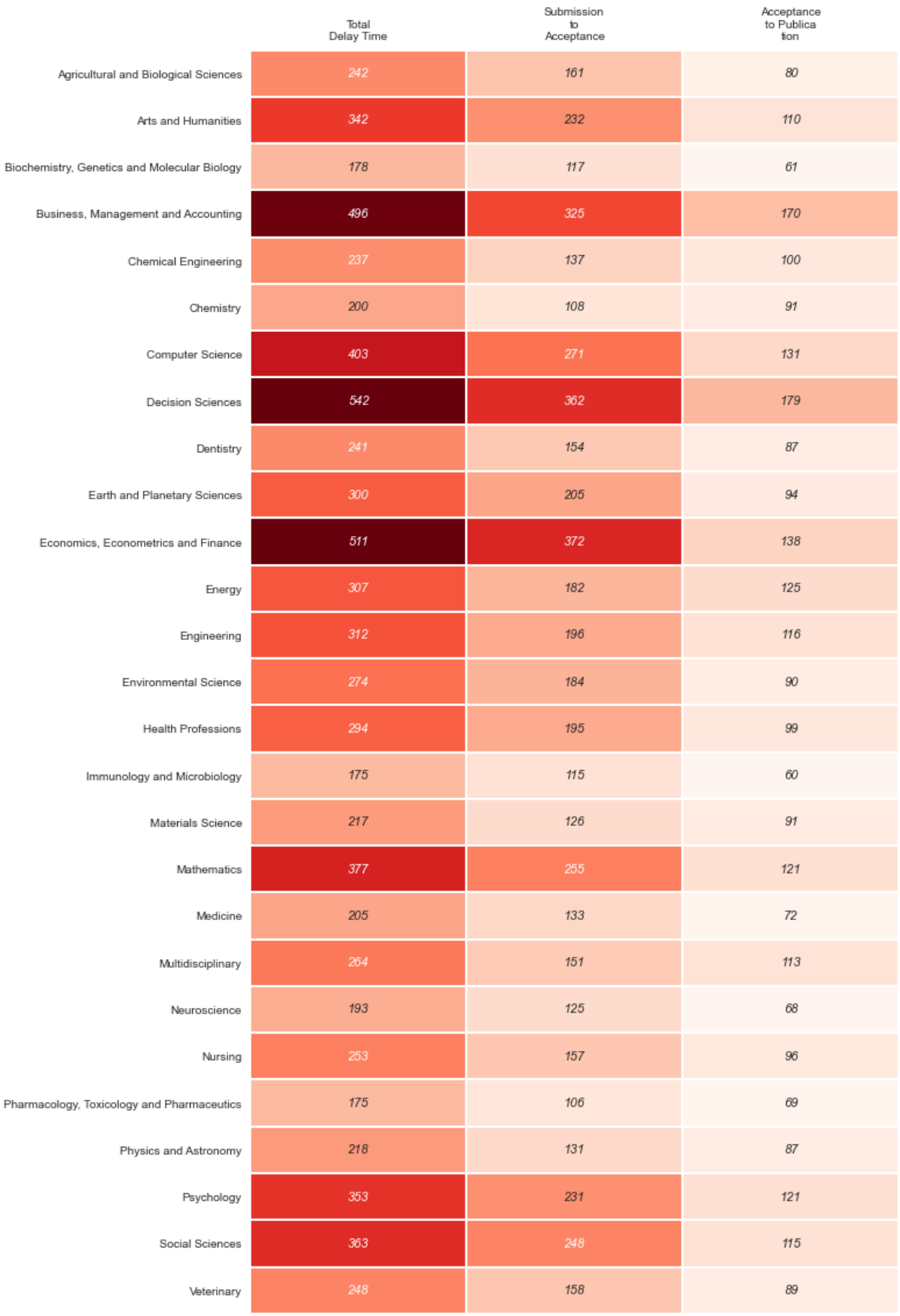
Environmental Science	586	587	588	589	590	590	590	591	593	593	594
Health Professions	553	556	557	560	561	562	561	563	563	566	569
Immunology and Microbiology	574	576	577	577	576	575	576	577	578	578	580
Materials Science	592	593	594	594	594	594	595	595	595	595	596
Mathematics	583	586	586	586	585	586	587	587	588	589	590
Medicine	599	599	599	599	599	599	599	599	599	599	599
Multidisciplinary	569	569	571	575	580	584	589	590	589	593	594
Neuroscience	577	578	579	579	579	580	581	581	581	581	580
Nursing	569	567	570	571	571	572	572	575	575	577	581
Pharmacology, Toxicology and Pharmaceutics	583	586	587	585	586	585	585	585	587	589	589
Physics and Astronomy	594	595	595	595	595	595	595	594	595	596	595
Psychology	566	569	571	576	577	577	578	580	581	583	582
Social Sciences	590	591	591	592	593	593	593	594	594	595	597
Veterinary	540	541	547	546	542	544	542	543	552	554	569

Appendix 2: Q1 journals - Required article sample size matrix

	2010	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**
Agricultural and Biological Sciences	635	635	662	697	652	708	680	706	680	677	698
Arts and Humanities	662	644	779	784	797	679	678	623	800	761	741
Biochemistry, Genetics and Molecular Biology	609	618	701	613	640	667	639	722	685	713	607
Business, Management and Accounting	684	585	682	591	673	589	639	638	627	702	622
Chemical Engineering	834	627	849	735	670	719	601	659	643	650	626
Chemistry	926	908	945	944	847	808	754	670	644	709	640
Computer Science	797	634	672	724	649	645	816	670	691	662	695
Decision Sciences	575	579	650	650	631	619	602	602	592	590	562

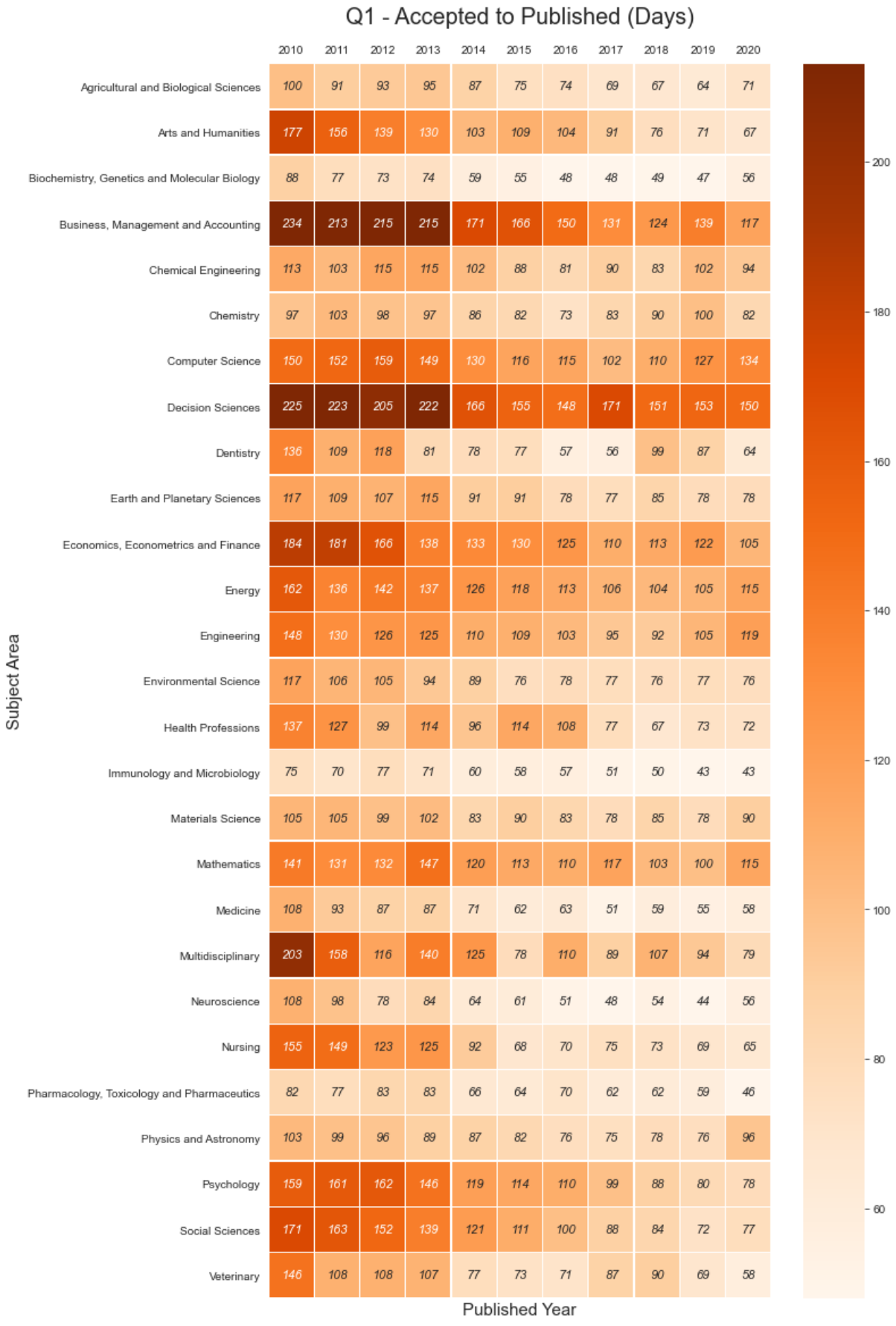
Dentistry	529	531	545	535	516	536	546	547	560	604	636
Earth and Planetary Sciences	878	926	938	925	814	632	816	773	761	760	742
Economics, Econometrics and Finance	688	647	687	613	644	629	608	595	595	580	638
Energy	803	773	786	799	787	801	792	671	625	660	600
Engineering	949	710	713	888	760	720	746	715	709	801	773
Environmental Science	782	722	834	805	653	635	611	612	630	624	662
Health Professions	555	578	613	586	627	587	562	576	576	570	573
Immunology and Microbiology	619	700	717	585	758	600	693	577	625	599	640
Materials Science	899	958	912	887	868	846	891	827	616	776	788
Mathematics	1004	996	1011	1082	1048	1024	1026	1070	973	931	859
Medicine	965	1030	1040	961	1068	979	1039	978	1036	1083	1006
Multidisciplinary	371	317	546	597	587	598	596	597	603	601	606
Neuroscience	721	624	769	670	655	680	628	630	632	677	710
Nursing	584	569	590	603	618	598	605	581	619	724	587
Pharmacology, Toxicology and Pharmaceutics	628	586	670	654	617	641	586	614	599	594	631
Physics and Astronomy	736	682	729	765	725	720	715	714	741	683	677
Psychology	732	752	684	617	615	638	646	595	663	645	609
Social Sciences	694	680	721	678	651	660	710	766	742	751	819
Veterinary	547	573	568	572	549	559	572	556	552	690	722

Appendix 3: Q1 journals – Collected article dataset matrix



Appendix 4: Mean average delay times heatmap

Appendix 5: Submitted to accepted average delay times heatmap



Appendix 6: Accepted to published average delay times heatmap

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação

Universidade Nova de Lisboa