Data description for 'No tickets for women in the COVID-19 race? A study on manuscript submissions and reviews in 2329 Elsevier journals during the pandemic'

Overview

This document provides an overview of the data used in the paper used for our paper *No tickets for women in the COVID-19 race? A study on manuscript submissions and reviews in 2329 Elsevier journals during the pandemic*. For further information see the paper itself.

The data used in the paper are organised in 8 datasets. All datasets are saved in *R* (RData) format.

```
## Authors.RData
## AuthorsCovid.RData
## AuthorsFirst.RData
## AuthorsQ1.RData
## AuthorsResearchPapers.RData
## AuthorsSolo.RData
## JournalAggregates.RData
## Referees.RData
```

Of these `Authors.RData` includes information about scientists acting as authors, `Referees.RData` information about scientist acting as referees, `JournalAggregates.RData` aggregate statistics per journal and week. All `Authors*.RData` datasets have the same structure of the `Authors.RData` one but focuses on specific subsets of submissions. All data are anonymized. The sections below present the structure of each dataset.

In addition, all data frames described below are available as separate *csv* files with names *'data frame name'.csv*. All files are included in the *AllDatasetsCsv.zip* archive.

Authors

The *Authors* dataset includes 4 data frames, one per research area. Each observation refers to a unique scientist and the dataset only includes scientists who submitted at least one paper in the corresponding research area in the period under consideration (week 6–22, 2018, 2019, and 2020).

Each of the 4 data frames has the same structure and only considers manuscripts submitted within a given research area in the periods under consideration. Research areas are *Health & medicine*, *Life sciences*, *Physical sciences & engineering*, and *Social sciences & economics* and the corresponding data frames are called `AuthorsHMS`, `AuthorsLS`, `AuthorsPS`, and `AuthorsSSE`.

Variables

All 4 data frames have the following structure:

- `ScientistID` Unique scientist ID

- `GenderRelaxed` Scientist gender according to the less restrictive version of the gender algorithm (hereafter "Gender1 algorithm", see the paper for details). Levels: {Men, Women}
- `GenderStrict` Scientist gender according to the more restrictive version of the gender algorithm (hereafter "Gender2 algorithm", see the paper for details). Levels: {Men, Women}
- `Seniority` Seniority variable based on salutation (see the paper for details). Levels: {not doctor, doctor, professor}
- `CountryCode` ISO Alpha-3 code of the scientist's affiliation country
- `Newsub2020` Number of submitted papers in the "covid" period between week 6 and week 22, 2020 (that is the sum of all newly submitted manuscripts in the period under consideration)
- `Newsub2019` Number of submitted papers in the period between week 6 and week 22, 2019 (that is the sum of all newly submitted manuscripts in the period under consideration)
- `Newsub2018` Number of submitted papers in the period between week 6 and week 22, 2018 (that is the sum of all newly submitted manuscripts in the period under consideration)
- `DeltaSub` Difference between the number of submitted papers in the "covid" period and the average of the corresponding periods in 2018 and 2019 (i.e, `newsub2020 - (newsub2019 + newsub2018) / 2`)
- `TimeSinceFirstPub` Time (in years) since the scientist's first publication recorded in Scopus

Subsets

These datasets are used to estimate models on specific subsets of the *Authors* dataset. All subsets have the same structure as the *Authors* dataset and result from additional filters applied when extracting the data from the Elsevier databases

Dataset name	Description
AuthorsCovid.RData	Only includes COVID-related submissions
AuthorsFirst.RData	Only includes submissions as first author
AuthorsSolo	Only includes submissions as solo author
AuthorsResearchPapers.RData	Only includes submissions of research papers
AuthorsQ1.RData	Only include submissions to top-quartile journals according to IF rankings

Referees

The *Referees* dataset includes 4 data frames (one per area), where each observation refers to a unique scientist. Only scientists who received at least one review invitation in the period under consideration (week 6-22, 2018, 2019, and 2020) are included.

Each of the 4 data frames has the same structure and only considers reviews within the corresponding research area in the period under consideration. The 4 data framed are called `RefereesHMS`, `RefereesLS`, `RefereesPS`, and `RefereesSSE`.

Variables

The 4 data frames have the following structure:

- `ScientistID` Unique scientist ID
- `GenderRelaxed` Scientist gender according to the Gender1 algorithm. Levels: {Men, Women}
- `GenderStrict` Scientist gender according to the Gender2 algorithm. Levels: {Men, Women}

- `Seniority` Seniority variable based on salutation. Levels: {not doctor, doctor, professor}
- `CountryCode` ISO Alpha-3 code of the scientist's affiliation country
- `Accepted2020` number of accepted review invitations in weeks 6–22, 2020
- `Declined2020` number of declined review invitations in weeks 6–22, 2020
- `Accepted2019` number of accepted review invitations in weeks 6–22, 2019

- `Declined2019` number of declined review invitations in weeks 6–22, 2019
- `Accepted2018` number of accepted review invitations in weeks 6–22, 2018
- `Declined2018` number of declined review invitations in weeks 6–22, 2018
- `PropAccepted2020` proportion of accepted reviews 2020 (`accepted2020 / (accepted2020 + declined2020)`)
- `PropAccepted2019` proportion of accepted reviews 2019 (`accepted2019 / (accepted2019 + declined2019)`)
- `PropAccepted2018` proportion of accepted reviews 2018 (`accepted2018 / (accepted2018 + declined2018)`)
- `DeltaProp` Difference in the proportion of accepted reviews between 2020 and the average of the two previous year (`propAccepted2020 - (propAccepted2019 + propAccepted2018) / 2`)
- `DeltaProp19.20` Difference in the proportion of accepted reviews between 2020 and t2019 (`PropAccepted2020 - PropAccepted2019`)
- `TimeSinceFirstPub` Time (in years) since the scientist's first publication recorded in Scopus

Journal aggregates

Journal aggregates includes two data frames — one for submissions and one for reviews — where each observation refers to the aggregate activity of one journal in one specific week of the whole period covered by the data (Jan 2018–May 2020). These data frames are only used for descritive statistics in the paper.

Submission data frame

The submission data frame is called `JournalSubmissions` and includes one observation per journal and week. For each journal, the total number of submissions of the week under consideration disaggregated by gender (using both algorithms) and seniority is reported for each journal. Note that this leads to multiple counts of the same paper if it has more than one author.

The first set of variables provides general information about the journal, date, and research area.

- `Journal` Unique Journal ID
- `Year` Submission year
- `Week` Submission week

- `Area` Submission area

The second set of variables groups submissions by seniority and gender based on the Gender1 algorithm:

- `G1FemaleNotDoctor` N. of papers submitted to the current Journal/Week by women with no title according to the Gender1 algorithm
- `G1FemaleDoctor` N. of papers submitted to the current Journal/Week by women with a Doctoral title according to the Gender1 algorithm
- `G1FemaleProfessor` N. of papers submitted to the current Journal/Week by women with a Professor title according to the Gender1 algorithm
- `G1FemaleUnknown` N. of papers submitted to the current Journal/Week by women with Unknown title according to the Gender1 algorithm

- `G1MaleNotDoctor` N. of papers submitted to the current Journal/Week by men with no title according to the Gender1 algorithm
- `G1MaleDoctor` N. of papers submitted to the current Journal/Week by men with a Doctoral title according to the Gender1 algorithm
- `G1MaleProfessor` N. of papers submitted to the current Journal/Week by men with a Professor title according to the Gender1 algorithm
- `G1MaleUnknown` N. of papers submitted to the current Journal/Week by men with Unknown title according to the Gender1 algorithm

- `G1UnknownNotDoctor` N. of papers submitted to the current Journal/Week by Unknown gender authors with no title according to the Gender1 algorithm
- `G1UnknownDoctor` N. of papers submitted to the current Journal/Week by Unknown gender authors with a Doctoral title according to the Gender1 algorithm
- `G1UnknownProfessor` N. of papers submitted to the current Journal/Week by Unknown gender authors with a Professor title according to the Gender1 algorithm
- `G1UnknownUnknown` N. of papers submitted to the current Journal/Week by Unknown gender authors with Unknown title according to the Gender1 algorithm

Variables in the third set are instead based on the Gender2 algorithm:

- `G2FemaleNotDoctor` N. of papers submitted to the current Journal/Week by women with no title according to the Gender2 algorithm
- `G2FemaleDoctor` N. of papers submitted to the current Journal/Week by women with a Doctoral title according to the Gender2 algorithm
- `G2FemaleProfessor` N. of papers submitted to the current Journal/Week by women with a Professor title according to the Gender2 algorithm
- `G2FemaleUnknown` N. of papers submitted to the current Journal/Week by women with Unknown title according to the Gender2 algorithm

- `G2MaleNotDoctor` N. of papers submitted to the current Journal/Week by men with no title according to the Gender2 algorithm
- `G2MaleDoctor` N. of papers submitted to the current Journal/Week by men with a Doctoral title according to the Gender2 algorithm
- `G2MaleProfessor` N. of papers submitted to the current Journal/Week by men with a Professor title according to the Gender2 algorithm
- `G2MaleUnknown` N. of papers submitted to the current Journal/Week by men with Unknown title according to the Gender2 algorithm

- `G2UnknownNotDoctor` N. of papers submitted to the current Journal/Week by Unknown gender authors with no title according to the Gender2 algorithm
- `G2UnknownDoctor` N. of papers submitted to the current Journal/Week by Unknown gender authors with a Doctoral title according to the Gender2 algorithm
- `G2UnknownProfessor` N. of papers submitted to the current Journal/Week by Unknown gender authors with a Professor title according to the Gender2 algorithm
- `G2UnknownUnknown` N. of papers submitted to the current Journal/Week by Unknown gender authors with Unknown title according to the Gender2 algorithm

Review data frame

The review data frame is called `JournalReviews` and includes one observation per journal and week. The total number of accepted and declined reviews of the week under consideration disaggregated by gender (using both algorithms) and seniority is reported for each journal.

The first set of variables provides general information about the journal, date, and research area.

- `Journal` Unique Journal ID
- `Year` Submission year
- `Week` Submission week

- `Area` Submission area

The second set of variables refers to accepted reviews with gender based on the Gender1 algorithm:

- `AcceptedG1FemaleNotDoctor` N. of accepted reviews in the current Journal/Week by women with no title according to the Gender1 algorithm
- `AcceptedG1FemaleDoctor` N. of accepted reviews in the current Journal/Week by women with a Doctoral title according to the Gender1 algorithm
- `AcceptedG1FemaleProfessor` N. of accepted reviews in the current Journal/Week by women with a Professor title according to the Gender1 algorithm
- `AcceptedG1FemaleUnknown` N. of accepted reviews in the current Journal/Week by women with Unknown title according to the Gender1 algorithm

- `AcceptedG1MaleNotDoctor` N. of accepted reviews in the current Journal/Week by men with no title according to the Gender1 algorithm
- `AcceptedG1MaleDoctor` N. of accepted reviews in the current Journal/Week by men with a Doctoral title according to the Gender1 algorithm
- `AcceptedG1MaleProfessor` N. of accepted reviews in the current Journal/Week by men with a Professor title according to the Gender1 algorithm
- `AcceptedG1MaleUnknown` N. of accepted reviews in the current Journal/Week by men with Unknown title according to the Gender1 algorithm

- `AcceptedG1UnknownNotDoctor` N. of accepted reviews in the current Journal/Week by Unknown gender authors with no title according to the Gender1 algorithm
- `AcceptedG1UnknownDoctor` N. of accepted reviews in the current Journal/Week by Unknown gender authors with a Doctoral title according to the Gender1 algorithm
- `AcceptedG1UnknownProfessor` N. of accepted reviews in the current Journal/Week by Unknown gender authors with a Professor title according to the Gender1 algorithm
- `AcceptedG1UnknownUnknown` N. of accepted reviews in the current Journal/Week by Unknown gender authors with Unknown title according to the Gender1 algorithm

The third set of variables refers to declined reviews with gender based on the Gender1 algorithm:

- `DeclinedG1FemaleNotDoctor` N. of declined reviews in the current Journal/Week by women with no title according to the Gender1 algorithm
- `DeclinedG1FemaleDoctor` N. of declined reviews in the current Journal/Week by women with a Doctoral title according to the Gender1 algorithm
- `DeclinedG1FemaleProfessor` N. of declined reviews in the current Journal/Week by women with a Professor title according to the Gender1 algorithm
- `DeclinedG1FemaleUnknown` N. of declined reviews in the current Journal/Week by women with Unknown title according to the Gender1 algorithm

- `DeclinedG1MaleNotDoctor` N. of declined reviews in the current Journal/Week by men with no title according to the Gender1 algorithm
- `DeclinedG1MaleDoctor` N. of declined reviews in the current Journal/Week by men with a Doctoral title according to the Gender1 algorithm
- `DeclinedG1MaleProfessor` N. of declined reviews in the current Journal/Week by men with a Professor title according to the Gender1 algorithm
- `DeclinedG1MaleUnknown` N. of declined reviews in the current Journal/Week by men with Unknown title according to the Gender1 algorithm

- `DeclinedG1UnknownNotDoctor` N. of declined reviews in the current Journal/Week by Unknown gender authors with no title according to the Gender1 algorithm
- `DeclinedG1UnknownDoctor` N. of declined reviews in the current Journal/Week by Unknown gender authors with a Doctoral title according to the Gender1 algorithm
- `DeclinedG1UnknownProfessor` N. of declined reviews in the current Journal/Week by Unknown gender authors with a Professor title according to the Gender1 algorithm
- `DeclinedG1UnknownUnknown` N. of declined reviews in the current Journal/Week by Unknown gender authors with Unknown title according to the Gender1 algorithm

The fourth set of variables refers to accepted reviews with gender based on the Gender2 algorithm:

- `AcceptedG2FemaleNotDoctor` N. of accepted reviews in the current Journal/Week by women with no title according to the Gender2 algorithm
- `AcceptedG2FemaleDoctor` N. of accepted reviews in the current Journal/Week by women with a Doctoral title according to the Gender2 algorithm
- `AcceptedG2FemaleProfessor` N. of accepted reviews in the current Journal/Week by women with a Professor title according to the Gender2 algorithm
- `AcceptedG2FemaleUnknown` N. of accepted reviews in the current Journal/Week by women with Unknown title according to the Gender2 algorithm

- `AcceptedG2MaleNotDoctor` N. of accepted reviews in the current Journal/Week by men with no title according to the Gender2 algorithm
- `AcceptedG2MaleDoctor` N. of accepted reviews in the current Journal/Week by men with a Doctoral title according to the Gender2 algorithm
- `AcceptedG2MaleProfessor` N. of accepted reviews in the current Journal/Week by men with a Professor title according to the Gender2 algorithm
- `AcceptedG2MaleUnknown` N. of accepted reviews in the current Journal/Week by men with Unknown title according to the Gender2 algorithm

- `AcceptedG2UnknownNotDoctor` N. of accepted reviews in the current Journal/Week by Unknown gender authors with no title according to the Gender2 algorithm
- `AcceptedG2UnknownDoctor` N. of accepted reviews in the current Journal/Week by Unknown gender authors with a Doctoral title according to the Gender2 algorithm
- `AcceptedG2UnknownProfessor` N. of accepted reviews in the current Journal/Week by Unknown gender authors with a Professor title according to the Gender2 algorithm
- `AcceptedG2UnknownUnknown` N. of accepted reviews in the current Journal/Week by Unknown gender authors with Unknown title according to the Gender2 algorithm

The fifth set of variables refers to declined reviews with gender based on the Gender2 algorithm:

- `DeclinedG2FemaleNotDoctor` N. of declined reviews in the current Journal/Week by women with no title according to the Gender2 algorithm

- `DeclinedG2FemaleDoctor` N. of declined reviews in the current Journal/Week by women with a Doctoral title according to the Gender2 algorithm
- `DeclinedG2FemaleProfessor` N. of declined reviews in the current Journal/Week by women with a Professor title according to the Gender2 algorithm
- `DeclinedG2FemaleUnknown` N. of declined reviews in the current Journal/Week by women with Unknown title according to the Gender2 algorithm

- `DeclinedG2MaleNotDoctor` N. of declined reviews in the current Journal/Week by men with no title according to the Gender2 algorithm
- `DeclinedG2MaleDoctor` N. of declined reviews in the current Journal/Week by men with a Doctoral title according to the Gender2 algorithm
- `DeclinedG2MaleProfessor` N. of declined reviews in the current Journal/Week by men with a Professor title according to the Gender2 algorithm
- `DeclinedG2MaleUnknown` N. of declined reviews in the current Journal/Week by men with Unknown title according to the Gender2 algorithm

- `DeclinedG2UnknownNotDoctor` N. of declined reviews in the current Journal/Week by Unknown gender authors with no title according to the Gender2 algorithm
- `DeclinedG2UnknownDoctor` N. of declined reviews in the current Journal/Week by Unknown gender authors with a Doctoral title according to the Gender2 algorithm
- `DeclinedG2UnknownProfessor` N. of declined reviews in the current Journal/Week by Unknown gender authors with a Professor title according to the Gender2 algorithm
- `DeclinedG2UnknownUnknown` N. of declined reviews in the current Journal/Week by Unknown gender authors with Unknown title according to the Gender2 algorithm