Gender Disparity in the Authorship of Biomedical Research Publications During the COVID-19 Pandemic: Retrospective Observational Study

Goran Muric[1], PhD; Kristina Lerman[1,2], PhD; Emilio Ferrara[1,2,3], PhD

Presented by Connor Fitzpatrick
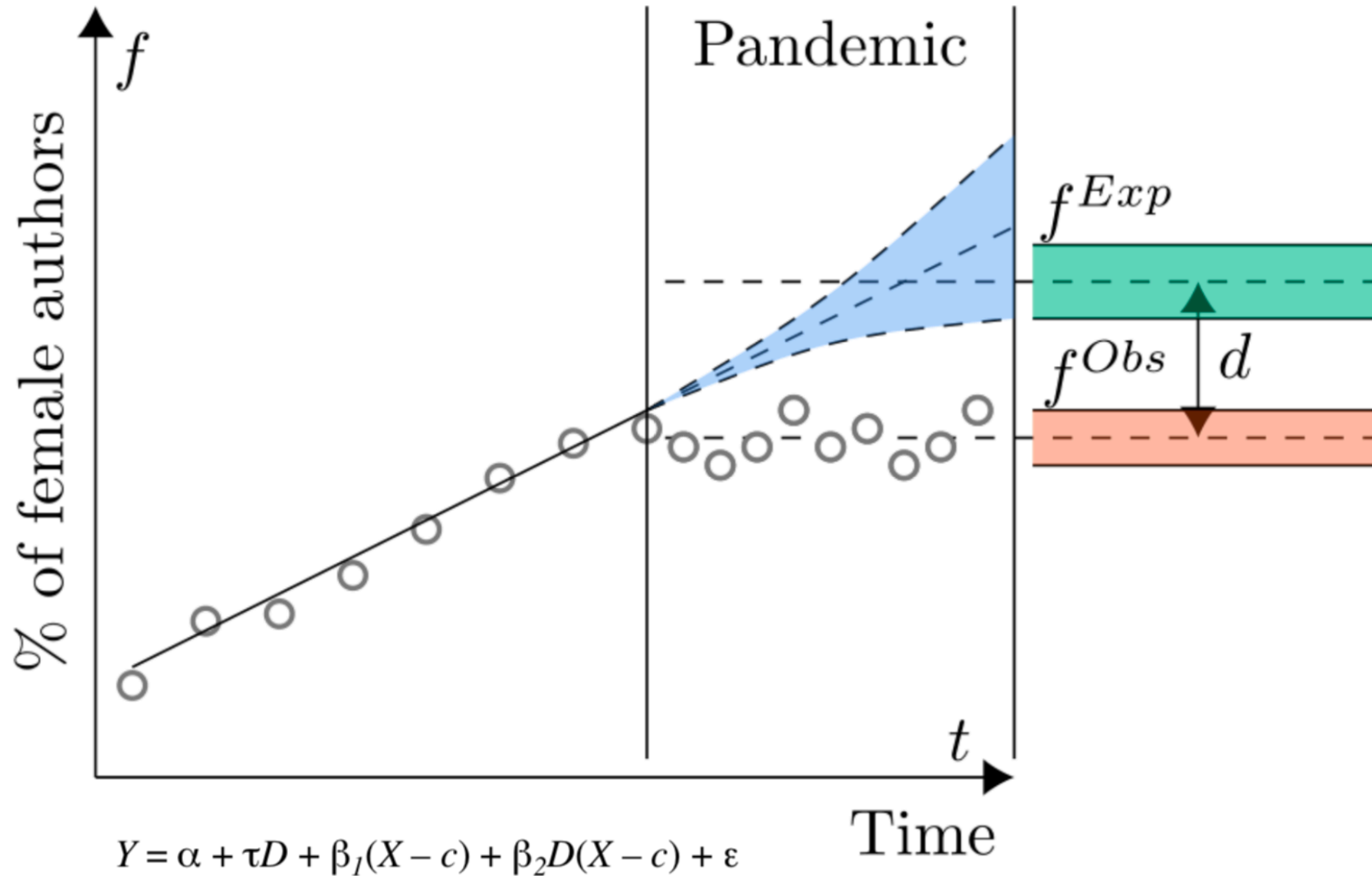WiSE Journal Club 7.13.21

Background & Hypothesis

- Female authorship has (generally) steadily increased in the biomedical industry in recent years

- Many socioeconomic impacts of COVID-19 have disproportionately affected women

- This paper aims to address whether this trend has carried over into biomedical publication rate & "productivity"

Methods

- Took 78,950 papers from Biorxiv, Medrxiv, & 62 selected Springer-Nature journals published by 346,354 unique authors
- Time frame of January 1st, 2019 – August 2nd, 2020
- Denoted March 15th, 2020 as the turning point of most shutdowns across the world
- Identified men & women identifying authors using the genderize.io API (61.4% men, 38.6% women)
  - Excluded low confidence name scores
- Used toponym of author's affiliate institution to identify country
- Compared expected proportions of publications to observed results using ordinary least squares (OSL) regression

Expected Model

$$Y = \alpha + \tau D + \beta_1(X - c) + \beta_2 D(X - c) + \varepsilon$$

Summary of All Results

Table 1. The expected and observed proportions of female authors disaggregated by the order of authorship and the topic.

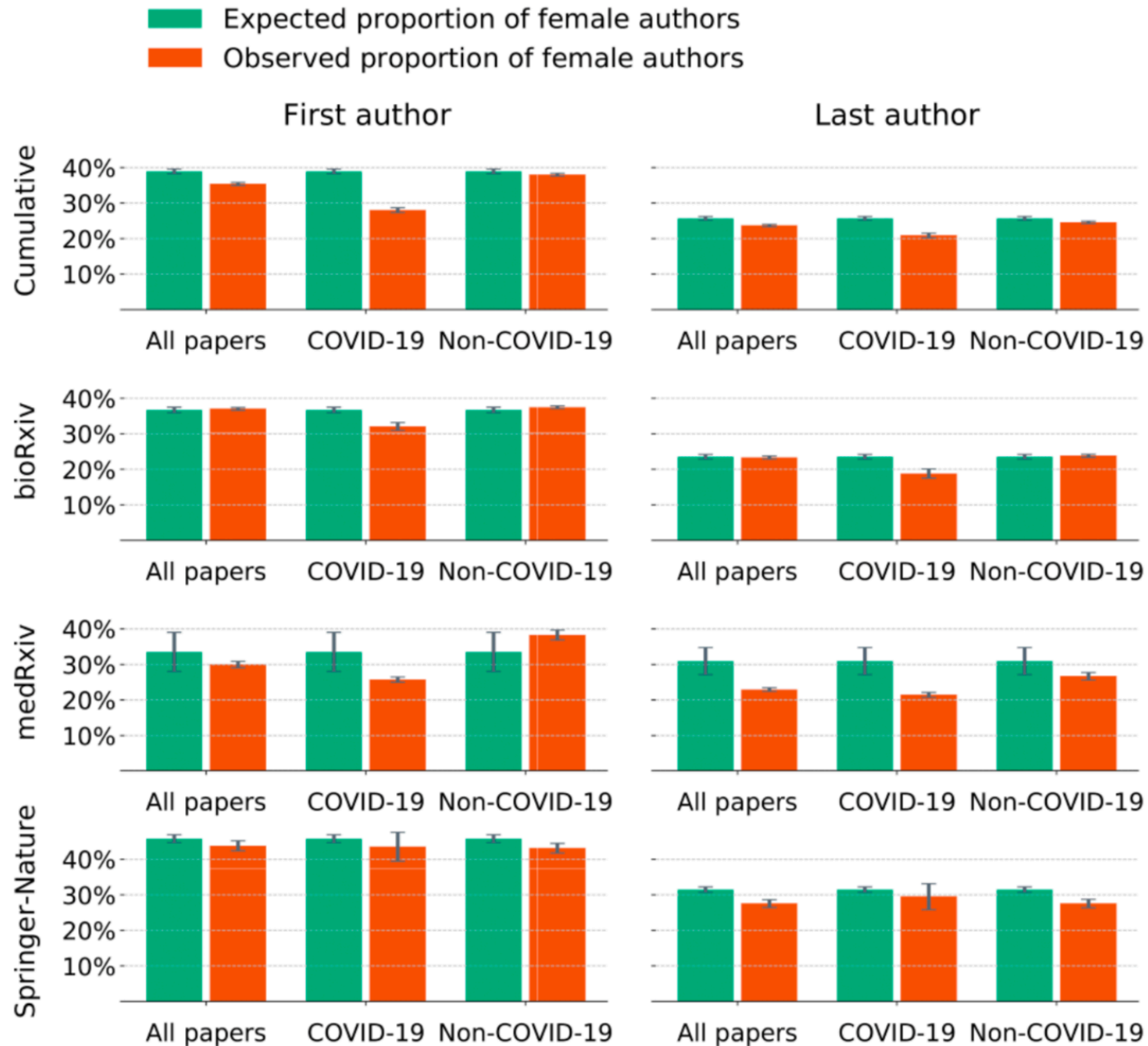| Author order and paper topic	Expected proportion		Observed proportion		Drop, %
	y_{est} [a]	S_{est} [b]	y [c]	σ_x [d]	
First					
All	0.389	0.007	0.353	0.004	9.142
COVID-19	0.389	0.007	0.280	0.007	28.031
Non-COVID-19	0.389	0.007	0.380	0.004	2.372
Last					
All	0.257	0.005	0.236	0.003	7.961
COVID-19	0.257	0.005	0.209	0.007	18.812
Non-COVID-19	0.257	0.005	0.246	0.003	4.416
Any					
All	0.354	0.003	0.348	0.002	1.578
COVID-19	0.354	0.003	0.341	0.009	3.530
Non-COVID-19	0.354	0.003	0.351	0.002	0.934
Solo					
All	0.210	0.030	0.137	0.008	34.586
COVID-19	0.210	0.030	0.137	0.023	34.514
Non-COVID-19	0.210	0.030	0.168	0.013	19.802

[a] y_{est}: the arithmetic mean of the estimate.

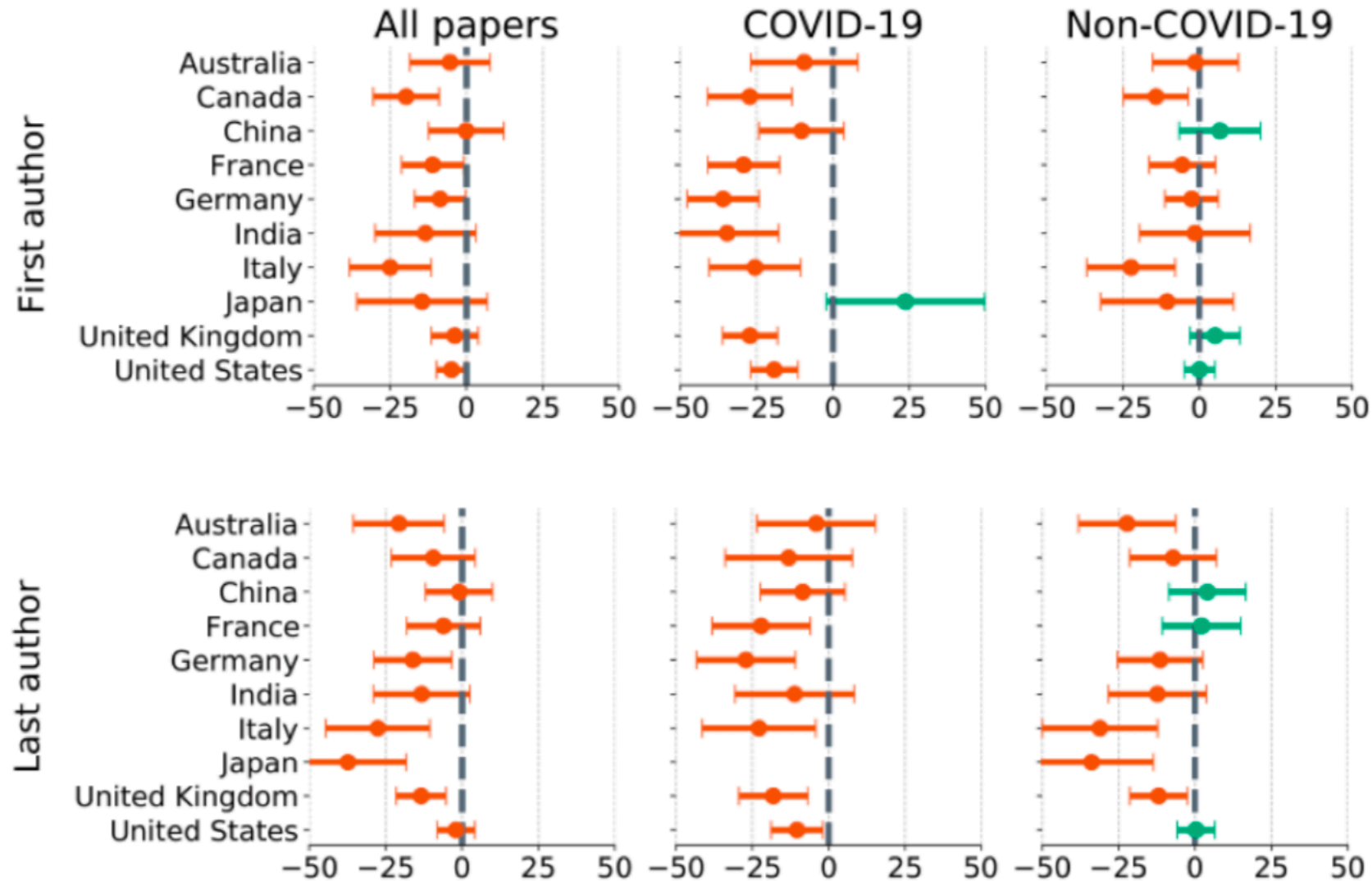[b] S_{est}: the mean standard error of the estimate.

[c] y: the arithmetic mean of the observation.

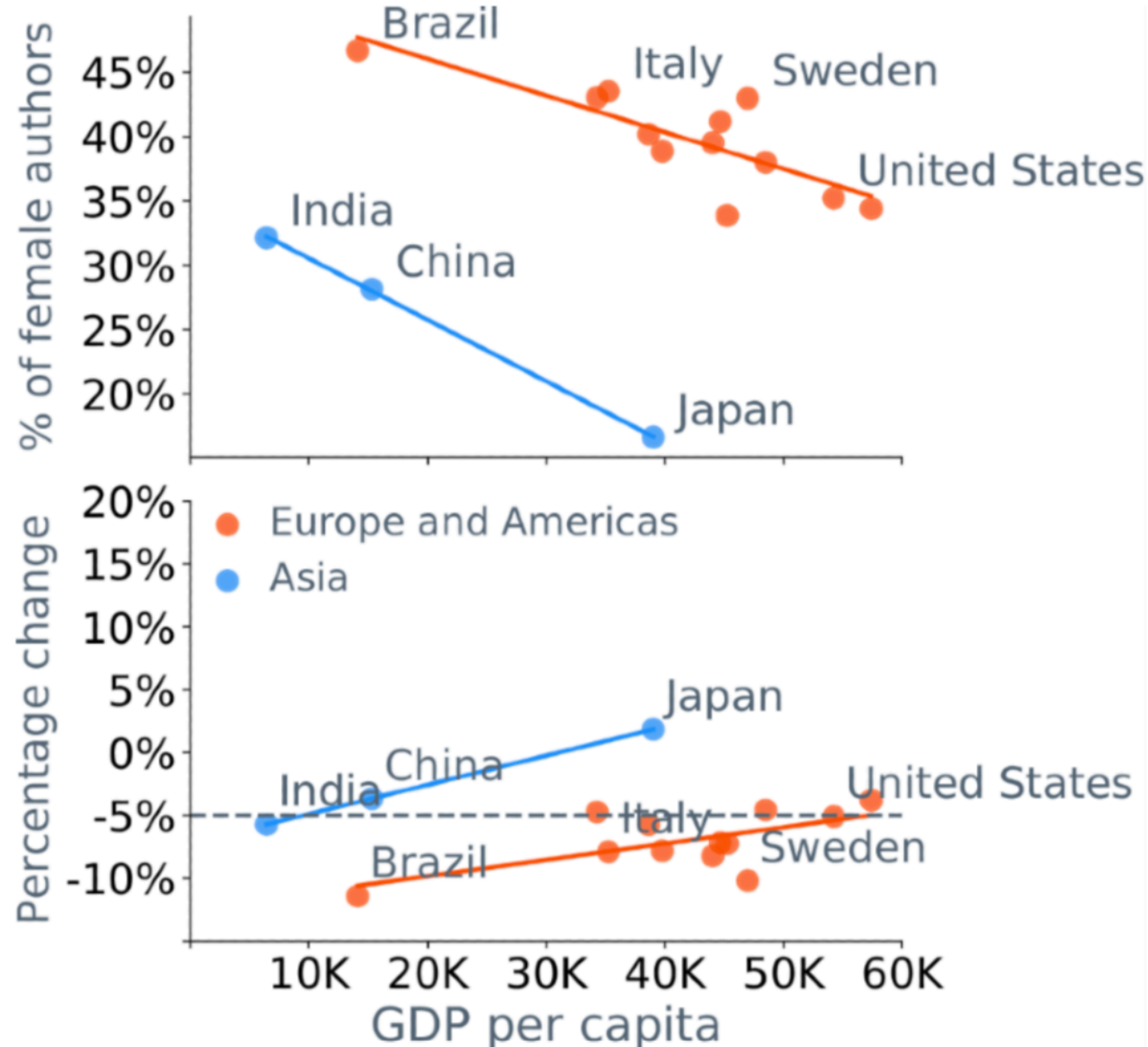[d] σ_x: the standard error of the mean of the observation.

Breakdown of Results

Results by Country

Results by Countries GDP/Capita

Strengths & Weaknesses

- ## Strengths
  - Large & (seemingly) robust dataset
  - Bunch of math & stats

- ## Weaknesses
  - Productivity =/= publication
  - Misgendering via genderize.io
  - Miscategorizing institution country

Feel free to contact me if you'd like to give a journal club in the future!