NOTEWORTHY

COUNTRY ANALYSIS OF PUBPEER ANNOTATED ARTICLES

14.08.2023

Just out of curiosity, after Scihub now an analysis of papers commented at the PubPeer website. Pubpeer is now also screened on a regular basis by Holden Thorp, the chief editor of Science...

Unfortunately I am loosing many records for incomplete or malformed addresses, while some preliminary conclusions can already be made when looking at my world map.

pubpeer.R grey indicates no data, black only a few, red numerous entries.

A further revision will need to include more addresses and also overall research output as a reference.

Other country level data are also interesting. Just to name a few

1/

Table 1. Association of Variables with Trust Score.

Variable (*n* of Respondents)		Median Trust Score (Interquartile Range)	*P* Value
Respondent Backgrounds			
Gender	Women (64)	2.33 (1.78 to 2.78)	Women vs Men 0.09
	Men (124)	2.56 (1.89 to 3.19)	
	Not disclosed (9)	2.44 (1.78 to 2.72)	
World region	Africa (6)	2.72 (2.42 to 3.58)	0.34
	Asia (21)	2.44 (2.11 to 3.39)	
	Europe (76)	2.44 (1.67 to 2.97)	
	North America (65)	2.33 (2.00 to 2.78)	
	South and Central America (8)	2.94 (1.92 to 3.44)	
	Oceania (14)	2.61 (1.97 to 3.36)	

Trust scores by continent https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10393470/pdf/rmmj-14-3-e0015.pdf

2/

Why do scientists falsify data? A matched-control analysis

Figure 3

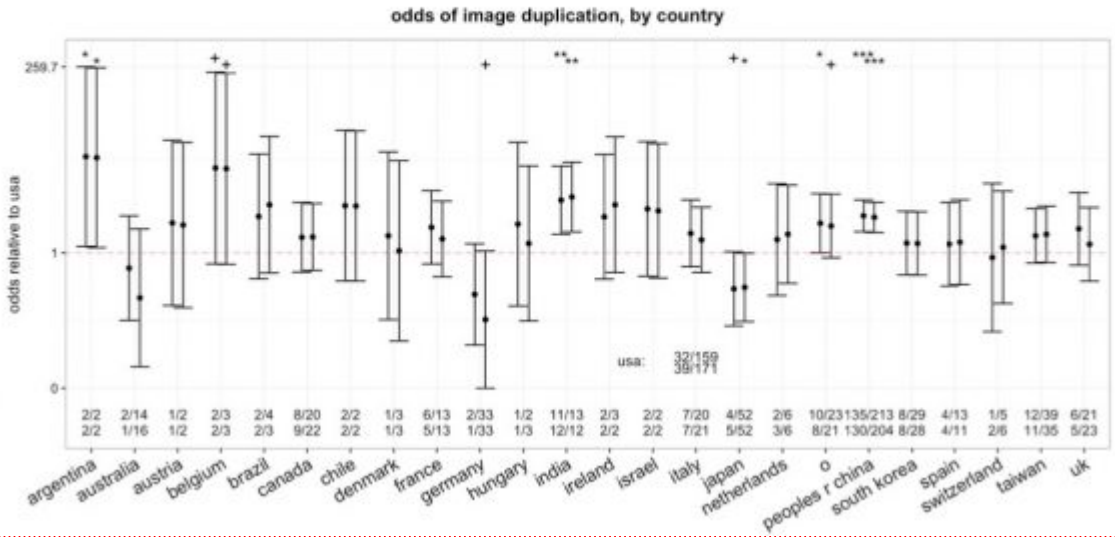


image duplication by country https://link.springer.com/article/10.1007/s11948-018-0023-7/figures/3

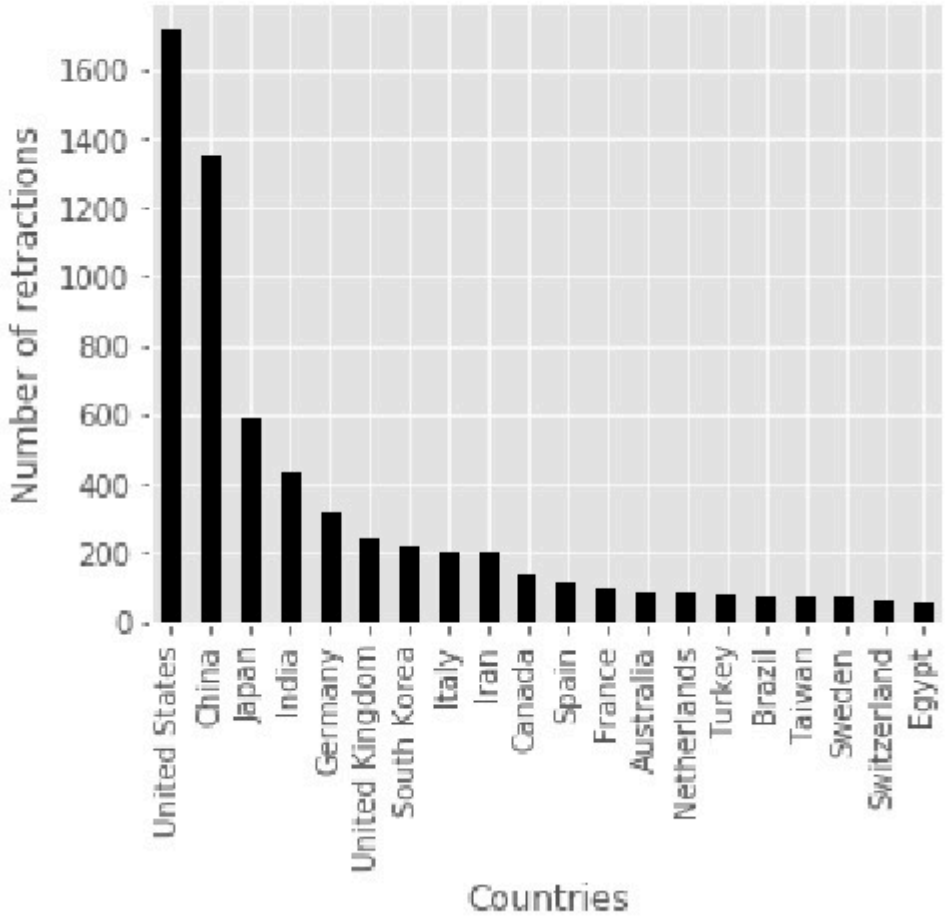
3/

Table 1 Nations with more than 50,000 documents from 2011 to 15 March 2017

Country	No. of documents 2011-current	No. of retracted articles	Misconduct ratio	Ranking based on misconduct ratio	Ranking based on no. of documents
China	2,741,274	4353	1.588E−03	1	2
Malaysia	157,198	50	3.181E−04	2	35
Mexico	121,193	31	2.558E−04	3	29
Taiwan	252,497	46	1.822E−04	4	17
Pakistan	71,350	10	1.402E−04	5	46
Iran	271,403	38	1.400E−04	6	22
Saudi Arabia	97,886	8	8.173E−05	7	44
Hong Kong	100,036	8	7.997E−05	8	31
South Korea	465,211	32	6.879E−05	9	12
Egypt	92,328	6	6.499E−05	10	42
India	747,844	39	5.215E−05	11	9
Singapore	117,089	6	5.124E−05	12	32
Thailand	78,124	4	5.120E−05	13	43
Australia	529,779	19	3.586E−05	14	11
Netherlands	343,352	12	3.495E−05	15	14
Romania	87,280	3	3.437E−05	16	41
Japan	787,157	27	3.430E−05	17	5
Canada	606,562	20	3.297E−05	18	7
Italy	624,340	18	2.883E−05	19	8
Greece	114,300	3	2.625E−05	20	27
United Kingdom	1,145,434	30	2.619E−05	21	3
Ireland	79,950	2	2.502E−05	22	38
Germany	1,010,967	25	2.473E−05	23	4
Czech Republic	130,262	3	2.303E−05	24	28
United States	3,876,791	88	2.270E−05	25	1
Portugal	134,433	3	2.232E−05	26	33
Austria	142,689	3	2.102E−05	27	24
Poland	238,095	5	2.100E−05	28	19
Belgium	192,437	4	2.079E−05	29	21

table 1 abbreviated https://link.springer.com/article/10.1007/s11948-017-9939-6

4/



retractions 2020 https://www.biorxiv.org/content/10.1101/2020.04.29.063016v6

5/

Country, No./No. (%)[a]	
China	1025/1 301 086 (0.08)
US	973/7 925 359 (0.01)
Japan	495/1 518 911 (0.03)
India	296/702 740 (0.04)
Germany	210/1 252 891 (0.02)
Italy	139/895 359 (0.02)
United Kingdom	136/2 045 725 (0.01)
Canada	61/1 006 743 (0.01)
Australia	51/733 657 (0.01)
France	41/835 361 (0.01)

Retractions
https://jamanetwork.com/journals/intemed/articlepdf/2779425/jamainternal_gaudino_2021_ld_210018_1627675018.88326.pdf

```
pubpeer[,"affiliation_list"] %>%
        (function(x) {
                strsplit(x,",") %>% sapply(., tail, 1) %>% unlist()
        }) %>%
        word(.,-1) %>%
        as_tibble() %>%
        mutate(value = gsub("[\\.]+", "",value)) %>%
        mutate(value = gsub("PR.*", "China",value)) %>%
        mutate( value = case_when
                (value== "" ~ NA,
                value== "States" ~ "USA",
                value== "University" ~ NA,
                value== "Republic" ~ NA,
                value== "ROC" ~ NA,
                value== "and" ~ NA,
                value== "Maryland" ~ "USA",
                value== "(mainland)" ~ "China",
                value== "Kong" ~ "HongKong",
                value== "Chemistry" ~ NA,
                value== "Engineering" ~ NA,
                value== "PAK" ~ "Pakistan",
                value== "Arabia" ~ "South Arabia",
                value== "Sciences" ~ NA,
                value== "Technology" ~ NA,
                value== "Medicine" ~ NA,
                value== "NY" ~ "USA",
                value== "America" ~ "USA",
                value== "York" ~ "USA",
                value== "Massachusetts" ~ "USA",
                value== "Hospital" ~ NA,
                value== "Zealand" ~ "New Zealand",
                value== "Pennsylvania" ~ "USA",
                value== "Africa" ~ "South Afria",
                .default = value)) %>%
        group_by(value) %>%
        count(value) %>%
        arrange( desc(n) ) %>%
        rename(region=value) %>%
        right_join(map_data('world'), by="region" ) %>%
        filter(region != "Antarctica") %>%
        ggplot() +
```

```
        geom_polygon(aes(long, lat, group = group, fill = n)) +
        coord_quickmap() +
        scale_fill_gradient(name = "N", trans = "log10", breaks =
c(10,100,1000), low = "black", high = "red", na.value = "lightgrey") +
        theme_void()
```