Neurorep exploratory analysis

Peder M. Isager

11/23/2020

Determining an initial set of candidates

The first step of our procedure was to determine a suitable set of candidate studies given our replication goals. Our research field of interest is social neuroscience, and our methodological interests pertain to fMRI research. We also determined to restrict our candidate set to studies published in the last ten years (later than 2009 at the time this decision was made). Our aim was thus to generate a representative sample of recently published fMRI studies within social neuroscience. In addition, we needed to determine a procedure for excluding studies from our candidate set that we would not be able to replicate (e.g. animal model research, highly invasive methodologies, research on patient groups, etc.).

We collected all records from the Web of Science database (citation). Because Web of Science does not have a predefined category of 'Social Neuroscience' we utilized two strategies for identifying social neuroscience research within the database. One strategy involved scraping all records from field-specific journals listed in the Web of Science. The other strategy involved scraping all records from Web of Science matching the key terms "social" and "fMRI". From the this initial set of records we then excluded a number of records when keyword information suggested the record would be unsuitable as a candidate in our replication effort.

Once a final set of candidate records had been determined, we explored the available bibliographic information to ensure that the sample indeed seemed representative of the field of social neuroscience fMRI research.

Methods/Procedure

We identified four journals in the Web of Science database as social neuroscience journals (*list journals*). Empirical articles published in these journals were identified by submitting the following search term to Web of Science:

[search term]

The search was conducted on YYYY-MM-DD. XXXX records were identified via this search strategy.

Searching field-specific journals is bound to miss many important studies in a field like social neuroscience, since many studies in this field are published in general topic journals like PLOS ONE, PNAS and Neuroimage. To be able to identify such studies and add them to our candidate set, we searched the entire Web of Science database for studies containing the keywords "social" and "fMRI" in either title or abstract. This general keyword combination is combatible with the description of many different topics in social neuroscience fMRI research, even for studies published in general topic journals.

Empirical articles containing the relevant keyword information were identified by submitting the following search term to Web of Science:

[search term]

The search was conducted on YYYY-MM.DD. XXXX records were identified via this search strategy.

Unsurprisingly, the two strategies yielded overlapping results, as studies published in social neuroscience journals are likely to contain the keywords "social" and "fMRI". After removing duplicate records, the two search strategies yielded XXXX unique empirical articles in total. These articles were considered our initial canidate set, and basic bibliometric information about each article, including author-provided keywords, were downloaded for all articles in the initial set.

Author PI and AV subsequently reviewed the 9807 unique author-provided keywords used to describe candidates in the initial set and curated a list of keywords to be used for further exclusion of articles. For example, we excluded all studies containing keywords such as “rats”, “canine”, “infants”, “als”, and any other term suggesting that the study would require access to a non-healthy/non-adult/non-human participant population, which would be unfeasible for our replication efforts. The complete records of excluded keywords can be found at [link to osf]. After excluding articles based on keyword information, our final set of candidates contained XXXX empirical articles.

Statistical analyses and exploration - summary

To verify that our final candidate set seemed representative of (human) social neuroscience research, we conducted several exploratory analyses of the rich bibliometric information available for each article via Web of Science. We explored the frequency distribution of journal outlets in order to verify that the journals most fequently chosen in our data correspond to popular publication outlets in social fMRI research. We explored the frequency distribution of Web of Science field categories (citation) to verify that categories such as “neurosciences”, “social psychology”, “psychology” and “multidisciplinary” were prevalent in our data.

In addition to exploring journal outlets and general field categories, we wanted to ensure that subfields and topics known to be prevalent in social fMRI research (e.g. social pain research [citation], face perception research [citation] and experimental paradigms from behavioral economy such as the dictator game [citation]). To this end, we acquired additional bibliometric information from the Centre for Science and Technology Studies (CWTS, [citation/link]) about prevalent citation clusters in our data (a proxy for scientific subfields contained within a larger research field). A citation cluster is determined by [ask Thed to write a short description on how CWTS determines citation clusters]. We analyzed the distribution of these clusters in our data, and we studied the frequency of category labels used to describe various clusters [ask Thed to write a summary of how these are derived]. Our goal was to verify that subfields and topics expected to be common were in fact freqently mentioned, and that no topic clearly irrelevant to social neuroscience were prominently featured.

To augment these analyses, we also utilized the statistical visialization software VOSviewer to extract commonly mentioned terms from the titles and abstracts of all studies, and we studied whether terms co-occured in line with our prior knowledge of terminology in different subfields of social neuroscience. All data included in the final dataset were subjected to analysis in VOSviewer with the parameters:

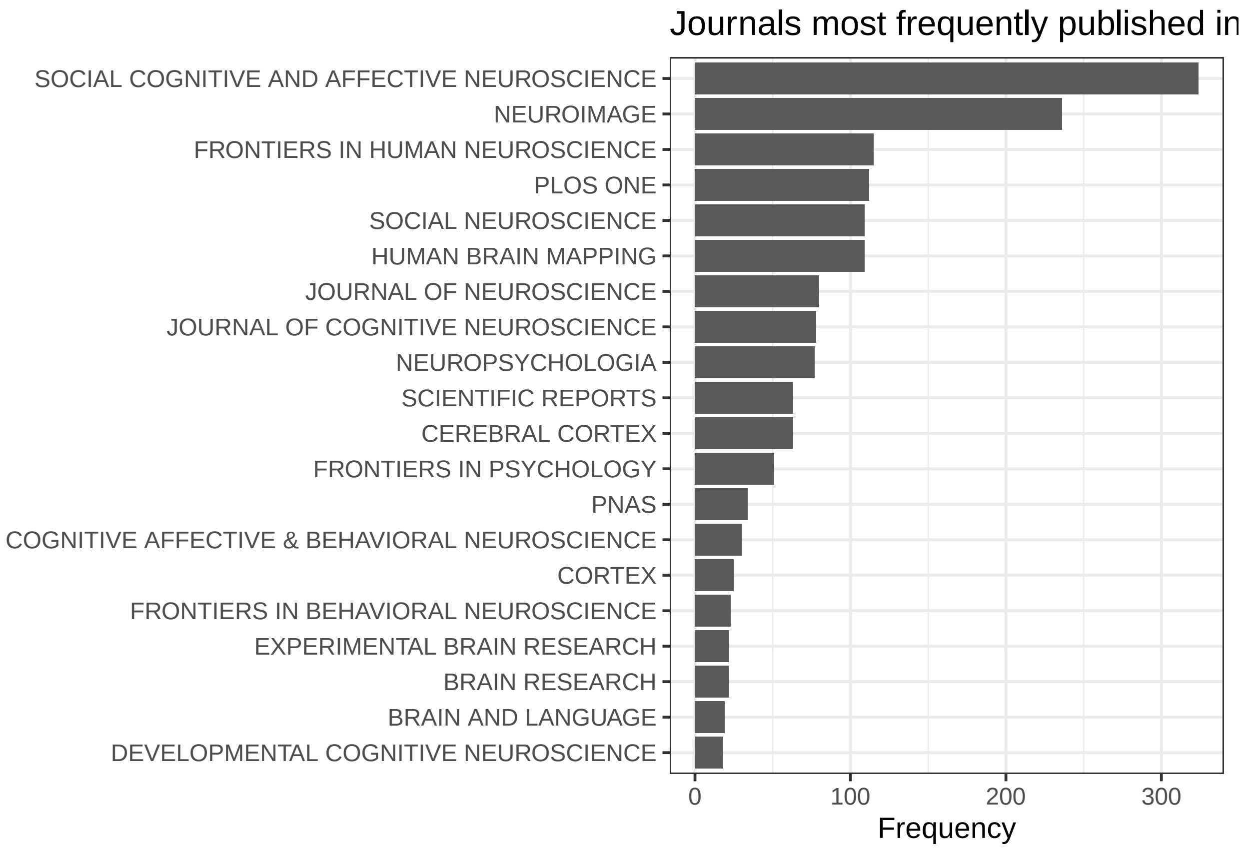
[list VOSviewer parameters and link to map files on OSF]

Results

Distribution of studies over journals

The records included in our dataset was published in 330 different journals. This is in line with our expectation that social neuroscience is a broad and loosely connected research field with a great number of subfield contained within.

Figure 1 displays the name and frequency of the 20 journals most frequently published in (*Peder draft note: We could also change this to be a table. Or we could turn the tables below into figures like this. Whatever helps readability the most.*). Unsurprisingly, two of the four journals from which all records were initially scraped were also among the most prominent journals in the final set of studies (Social Cognitive and Affective Neuroscience, and Social Neuroscience). Besides these two, the sample appears to be dominated by journals that are either general topic, (Plos ONE and PNAS) or general neuroscience/psychology (e.g. Neuroimage, Frontiers Psychology, Cortex). The lack of specialist journals in the top end of the frequency distribution is likely due to the fact that these journals only serve a smaller subsection of the larger community of social neuroscientists, while journals like Neuroimage and Plos ONE can, in principle, serve them all.



Distribution of studies over Web of Science categories

The records in our dataset was classified as being members of 178 unique Web of Science categories. Table 1 displays the name and frequency of the 20 Web of Science categories most frequently tagged.

Table 1: Journals most frequently published in

field	frequency
Neurosciences; Neuroimaging; Radiology, Nuclear Medicine & Medical Imaging	345
Neurosciences; Psychology; Psychology, Experimental	333
Neurosciences	291
Neurosciences; Psychology	225
Multidisciplinary Sciences	222
Behavioral Sciences; Neurosciences	112
Neurosciences; Psychology, Experimental	97
Psychology, Multidisciplinary	81
Behavioral Sciences; Neurosciences; Psychology, Experimental	77
Psychology, Experimental	38
Psychology, Social	27
Audiology & Speech-Language Pathology; Linguistics; Neurosciences; Psychology, Experimental	19
Psychology, Developmental; Neurosciences	18
Endocrinology & Metabolism; Neurosciences; Psychiatry	13
Psychiatry	13
Psychology, Biological; Neurosciences; Physiology; Psychology; Psychology, Experimental	12
Anatomy & Morphology; Neurosciences	10

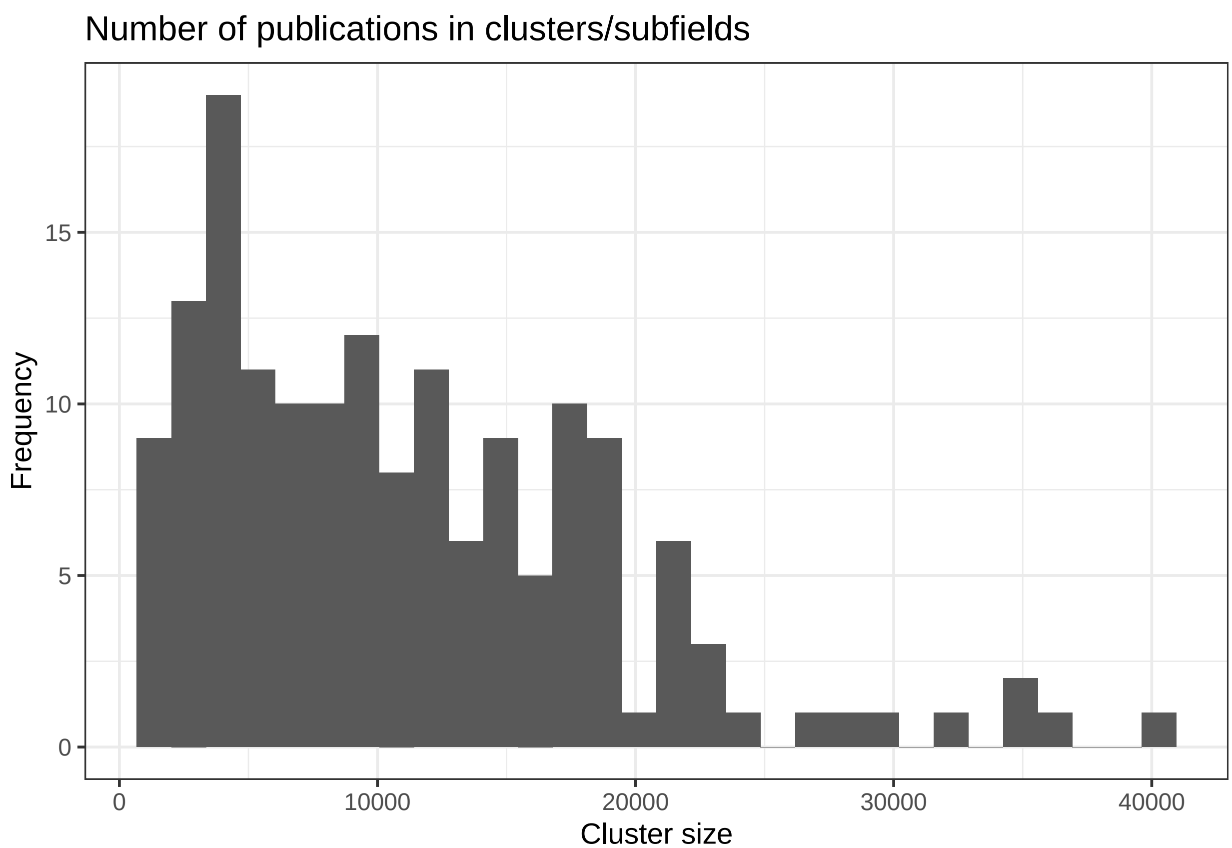
field	frequency
Neuroimaging	10
Neurosciences; Pharmacology & Pharmacy; Psychiatry	10
Neurosciences; Physiology	9

Citation clusters and frequently co-occuring keywords

Examining bibliometric information from CWTS, we found that the records in our dataset is contained in 162 unique citation citation clusters. As shown in Figure 2, the number of articles in each cluster varies substantially (min=829, median=1.2354×10^4, max=3.977×10^4).

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## Warning: Removed 1 rows containing non-finite values (stat_bin).
```



To better understand the scientific topic covered by these citation clusters, we inspected the category labels assigned to each cluster by CWTS. In total, the citation clusters were associated with 774 unique labels. Table 2 displays the frequency of the 50 most frequently mentioned category labels in our data.

Table 2: Most frequent cluster labels

label	frequency
intertemporal choice	364
decision making	358
delay discounting	358
impulsivity	358

label	frequency
iowa gambling task	358
imitation	318
action observation	258
empathy	258
mirror neuron	258
motor imagery	258
attentional bias	210
fear	207
emotional face	203
facial expression	203
social anxiety	203
default mode network	180
fmri	180
fmri data	180
functional connectivity	180
resting state	180
alzheimer	125
face processing	118
face recognition	118
facial identity	118
prosopagnosia	118
unfamiliar face	118
n400	109
primary progressive aphasia	109
semantic dementia	109
visual word recognition	109
contamination	67
disgust	67
disgust sensitivity	67
moral dilemma	67
moral judgment	67
death anxiety	65
mind	65
mortality salience	65
ostracism	65
social exclusion	65
terror management	65
autobiographical memory	63
expressive writing	63
generativity	63
mental time travel	63
rumination	63
false belief	60
infant	60
month old infant	60
effect	52

To complement the cluster information from CWTS, we utilized the VOSviewer analysis tool to extract topic-related keywords from article titles and abstracts, and analyze co-occurences between these keywords. Figure 3 displays the co-occurence map between commonly mentioned keywords in our dataset.

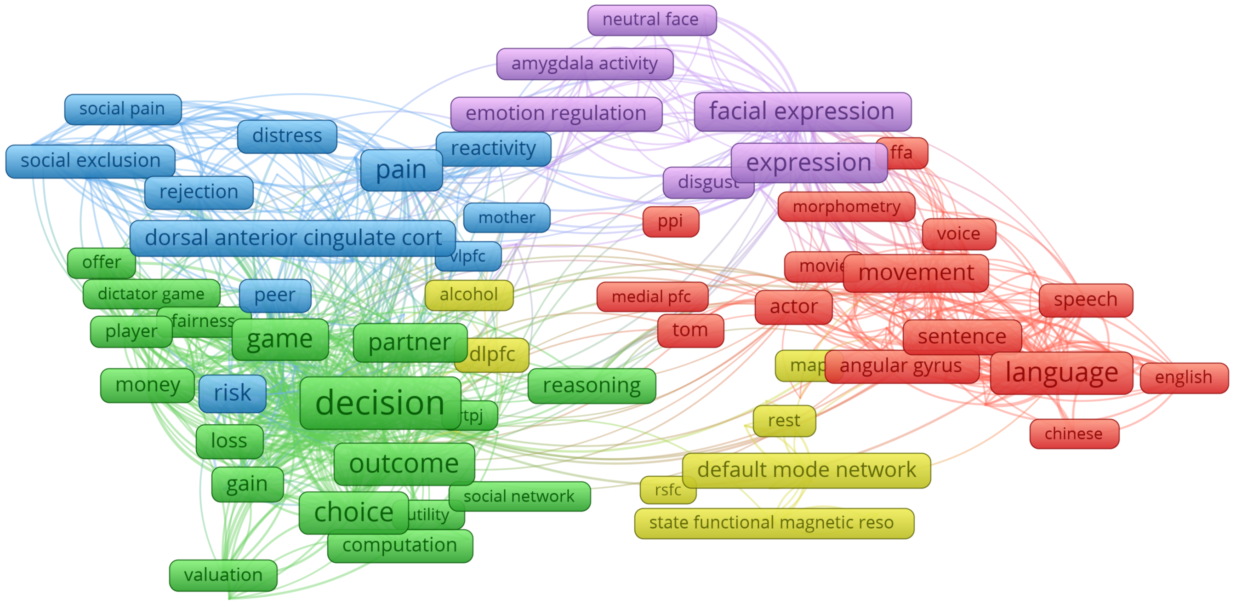


Figure 1: Figure 3: VOSviewer map of title/abstract keyword co-occurences

Discussion

Based on the bibliometric information summarized above, we feel confident that we have successfully managed to sample articles from human social fMRI research. Journals common in the field of social neuroscience are also frequent within our data. The Web of Science category distribution is similarly consistent with what we would expect from studies sampled from social neuroscience research, with categories such as "Neurosciences; Psychology; Psychology, Experimental" and "Multidisciplinary Sciences" being among the most common. On the other hand, it is somewhat surprising that categories such as "Psychology, Social" and "Neuroimaging" are not more prevalent in a dataset that is supposed to contain fMRI studies of social psyhological phenomena.

The concern about prevalence of fMRI methodology and social psychology phenomena in the dataset is however relieved by inspecting the distribution of CWTS cluster labels and the VOSviewer co-occurence map. On the one hand, "fmri" and "fmri data" are among the 50 most common labels used to describe citation clusters to which our data belongs. On the other hand, terms such as "imitation", "empathy" "mirror neuron", "facial expression", and "social exclusion" suggests that topics common in social fMRI research are also well-represented in our dataset. The VOSviewer co-occurence map shows that topics frequent in article titles and abstract overlap with topics frequent within the CWTS cluster labels.

The co-occurence map also suggests a number of larger subtopics within the data. As expected from a set of articles sampled from social neuroscience, language, social pain and exclusion, and face perception seem to be highly prevalent themes. Not consistent with our expectations is the prominent cluster of studies related to the default mode network and functional connectivity. Visual inspection of titles that are categorized in the "default mode" cluster by CWTS suggests that many of these articles are purely methodological, and a vast majority do not seem to be concerned with social neuroscience as such.

Another unexpectedly prevalent topic in the co-occurence map is that centered around decision-making. Convergently, the 5 most frequent CWTS cluster labels (table 2) all seem related to choice and decision making, which is not obviously a topic sorted under social neuroscience. Reviewing the titles and abstracts of articles within the CWTS "decision making" cluster, reveals a more nuanced picture. The citation cluster described by the labels "intertemporal choice", "decision making", "delay discounting", "impulsivity" and "iowa gambling task" is the most prevalent cluster in our data (358 articles in our data belong in this cluster). However, the CWTS labels used to describe this cluster are not necessarily representative of the articles from this cluster that are included our dataset. For example, although “iowa gambling task” is descriptive of the cluster as a whole, only a single article from this cluster in our dataset even mentions the Iowa gambling task. We therefore consider it likely that we have sampled a biased subset of articles from this cluster, which seems plausible considering that the cluster contains a total of 1.3168×10^4 articles. The articles from this cluster that are contained in our data concern a variety of topics, most of which more clearly related to social psychology than the cluster labels would indicate. For example, neuromarketing designs and study designs common in behavioral economy (e.g. ultimatum and trust games) appear frequently, which also explains the frequent co-occurences of terms like “decision”, “outcome”, “choice”, “partner” and “game” (Figure 3). However, we should note that there also appears to be a number of purely methodological articles in this subset, suggesting that our method of excluding methodological articles by article keyword information was not entirely successful.

In summary, our exploratory analyses suggest that we have been largely successful in curating a large set of studies from the social neuroscience literature that employ fMRI methodology and otherwise adhere to our inclusion criteria. However, we remind the reader that the results above summarizes only a subset of a larger collection of bibliometric information available for our dataset. The results we report are those we believe are most relevant for evaluating whether we have successfully sampled the population of human social fMRI research. However, the full dataset including all bibliometric variables are available at [OSF link to data] for the curious/sceptical reader.

Operationalizing value and uncertainty

Having determined on a set of candidate articles to consider for replication, the next step in our selection procedure was to derive a quantitative estimate of replication value for each replication candidate included in our dataset. In theory, this simply involves determining a suitable formula for estimating RV, collecting the necessary data for each candidate, and applying the formula to each candidate study in the dataset. However, in practice there are several additional challenges to consider.

First, we must settle on a quantitative definition of RV that is likely to be valid for estimating the expected utility of our replication attempt (Isager et al. 2020). We determined to use the formula described in Isager et al. (2020 - thesis chapter 2) as our primary definition of RV. However, this formula is not yet validated empirically, neither in general nor in social fMRI research specifically. Thus, in addition to collecting the information necessary to calculate the formula described in Isager et al. (2020 - thesis chapter 2), we aimed to identify additional quantitative indicators that might be important for estimating RV. We also aimed to collect quantitative information that would let us compare the performance of the Isager et al. (2020) indicator with other potential operationalizations of RV (e.g. Field et al. 2019, which required information about bayes factors).

Second, given that the target of a replication study is a claim (Isager et al. 2020 - chapter 1), and given that any article in our dataset may contain multiple claims, we must decide which claims from each article to focus our formula RV estimates on. We initially determined to focus our efforts on the main claim from each study from each article in our set of candidates. This means that each article in our dataset actually represents as many replication canidates as there are empirical studies reported in that article. We subsequently began the process of coding, for each individual study in each article, the main finding reported for that study.

Third, we needed to determine which quantitative indicators of “value” and “uncertainty” are feasible to collect in practice, as this would determine which operationalizations of RV we could consider estimating. For instance, we knew that the formula of Isager et al. (2020) ideally requires enough statistical information that a standard error can be calculated. This implies that it must be possible for us to identify statistical tests of each claim under consideration, and also that the necessary information about standard deviations, sample size etc. must be available for each of these tests. Finally, given the large number of canidates we are considering, we reuire a quick and efficient method for collecting the necessary quantitative information.

Operationalizing "value" - methods/procedure

We utilized various citation impact metrics as indicators of the value of each replication candiate, following the equation and rationale laid out in Isager et al. (2020, thesis chapter 3). We needed to select a single bibliometric source to rely in for citation impact estimates. However, in practice their are sources to choose from (Crossref, Scopus, Web of Science, etc.), and no principled reason for preferring one over the other. We therefore decided to collect citation count information from several bibliometric sources and inspect the similarity of the citation count estimates provided. We collected citation count data from Web of Science (provided with the bibliometric data collected when identifying the initial canidate set), Crossref (using the rcrossref package in R [citation]), Scopus (using the rscopus package in R [citation]), and CWTS (provided by CWTS staff).

To address the fact that different subfields may have different citation practices that inflate citation counts in some fields compared to others, we also collected field-normalized citation scores from the CWTS database (see [citation] for details about the normalization procedure). Since it is not completely clear whether field-normalized citation scores should be preferred to non-normalized scores for calculating replication value (Isager et al. 2020 - either chapter 1 or 3 discusses whether normalizing scores makes sense) our initial goal was simply to observe the correlation between field-normalized and non-normalized scores, to better understand the impact of choosing one or the other.

Finally, we also collected Altmetric scores as an alternative operationalization of impact. Altmetric scores are known to be only weakly associated with more traditional citation metrics (citation), and presumably captures

Operationalizing "uncertainty" - methods/procedure

Consulting field experts to identify potential quantitative indicators of interest

Identifying the "main claim/finding" for each article

```
cor.dat <- select(.data = data.all,
       TC_2020,
       TC,
       crossref_citations,
       scopus_citations,
       tcs,
       tncs,
       altmetric_score,
       sample_size,
       years.since.pub)
kable(x = describe(cor.dat))
```

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
TC_2020	1	1358	30.321794	41.033933	16.00	21.625911	16.308600	0.00	487.00	487.00	3.4621437	19.3344227	1.1135078
TC	2	1358	21.191453	32.001545	10.00	14.304228	13.343400	0.00	416.00	416.00	3.7443832	24.8510422	0.8684025
crossref_citations	3	1350	31.368148	42.009171	17.00	22.452778	17.791200	0.00	484.00	484.00	3.3919042	17.9934619	1.1433450
scopus_citations	4	1343	32.960536	44.353939	27.00	23.518140	17.791200	0.00	506.00	506.00	3.3735102	17.6966051	1.2103032
tcs	5	1332	20.421173	30.449461	10.00	13.797373	11.860800	0.00	388.00	388.00	3.7920402	24.5144185	0.8343101
tncs	6	1332	1.294084	1.396575	0.84	1.050047	0.830256	0.00	11.16	11.16	2.4477129	9.1268152	0.0382659
altmetric_score	7	1156	33.593531	105.261453	3.35	11.472126	5.819960	0.25	1195.48	1195.23	7.3918476	68.0524153	3.0959245
sample_size	8	1358	42.315169	94.340675	24.00	27.799632	11.860800	1.00	1553.00	1552.00	10.2018133	125.3946923	2.5600538
years.since.pub	9	1358	5.798233	2.687167	6.00	5.693015	2.965200	1.00	11.00	10.00	0.2677148	-0.9607254	0.0729197

Results

Sample size inter-rater reliability

```
## Percentage exact agreement
all.match <- sum(data.irr$matches_all)/nrow(data.irr)
orig.BA.match <- sum(data.irr$matches_orig_BA)/nrow(data.irr)
orig.PhD.match <- sum(data.irr$matches_orig_PhD)/nrow(data.irr)
BA.PhD.match <- sum(data.irr$matches_BA_PhD)/nrow(data.irr)

icc <- ICC(data.irr[, c("sample_size_orig", "sample_size_BA", "sample_size_PhD")])
```

```
## boundary (singular) fit: see ?isSingular
```

```
g.irr <- ggplot(data = data.irr, aes(x=sample_size_orig)) +
  geom_point(aes(y=sample_size_BA), col = "#E69F00", alpha = 0.6) +
  geom_point(aes(y=sample_size_PhD), col = "#56B4E9", alpha = 0.6) +
  theme_bw() +
  scale_color_manual(values=c("#E69F00", "#56B4E9"), name = "Coder", labels = c("BA", "PhD")) +
  scale_x_continuous(trans='log10') +
  scale_y_continuous(trans='log10')
```

Overall, there was a high, albeit far from perfect, agreement between the three coders (percentage exact agreement = 0.772). The BA double coder and the PhD coder had a slightly higher agreement rate (percentage exact agreement = 0.828) than either one had with the original BA coders (percentage exact agreement between original BA coders and BA double coder = 0.816, percentage exact agreement between original BA coders and PhD double coder = 0.828). The intraclass correlation coefficient between raters was high, ICC = 0.8245799, CI95%[0.7951546, 0.8511194].

Figure X displays

Calculating replication value