Bibliometric Analysis of Resting State fMRI Literature

Matthew K. Doherty[1], Ayesha Anwar[1], Sam Barberie[1], Caitlin Hinz[1], Michelle Kaplan[1], Anna Rachlin[1], Michael P. Milham[1,1], Stephen M. LaConte[1,1], R. Cameron Craddock[1,1,]

[a] *Child Mind Institute, New York, New York*
[b] *Nathan Kline Institute for Psychiatric Research, Orangeburg, New York*
[c] *Virginia Tech Carilion Research Institute, Roanoke, Virginia*
[d] *School of Biomedical Engineering and Sciences, Virginia Tech, Blacksburg, Virginia*

Abstract

Some cool stuff about why the Child Mind Librarian is so freaking cool ... and why we care, and other things of relative importance.

Keywords: Functional connectivity, connectome, group study, effective connectivity, fMRI, resting-state

1. Introduction

Since its initial observation in 1995 [**?**], resting state functional magnetic resonance imaging (R-fMRI) has exploded in popularity as a technique for studying the brains functional architecture. Although initially plagued by controversies related to the physiological underpinnings and unconstrained nature of R-fMRI [**?** **?**], it has persevered to become a standard assessment in basic, clinical and cognitive neurosciences [**?** **?**]. The result is a burgeoning literature dedicated to R-fMRI that is quickly becoming unwieldy for researchers to navigate. The Child Mind Institute (CMI) Librarian initiative addresses this challenge by providing a hand-curated reference library of R-fMRI literature (http://www.mendeley.com/profiles/cmi-librarian). This library consists of 1,721 publications (as of December 24, 2012) and requisite metadata to facilitate the systematic review of its contents. We inaugurate the availability of this open resource by performing a bibliometric analysis to assess the current state of R-fMRI literature.

Bibliometrics involves the application of mathematics and statistical methods to a body of literature to illuminate its course of development and measure its impact [**?**]. Although primarily concerned with analyzing citations to estimate the impact of researchers and particular papers within the field, bibliometrics can also identify common themes within in the literature. Bibliometric analyses differ from more standard literature reviews in that they are largely automated and may overlook or misinterpret details of individual publications that would be apparent to a human reader. On the other hand, bibliometric analyses are data-driven and can be more comprehensive, using a larger fraction of the corpus to support its conclusions. The quality of the analyses is determined by the quality of the literature database used for the analysis. The CMI Librarian resting state literature database is ideal for automated data mining because it has been hand-vetted to remove irrelevant publications, and to provide basic keywords to inform the analysis.

This automated statistical analysis of R-fMRI literature incorporated information from the CMI librarian database, PubMed (http://pubmed.gov), and the full text of the publications. Information across these resources

*Corresponding author: cameron.craddock@childmind.org

was combined to gain insight into publication growth, venues and patterns. Citation analyses were performed to identify publications and researchers with the greatest impact, as well as to identify working groups of researchers who tend to co-publish. Key word analyses were performed to identify experimental methods, cognitive domains, clinical disorders, and brain regions most commonly discussed in the literature. Together the results from these analyses narrate the history of R-fMRI.

2. Methods

2.1. CMI Librarian

The CMI Librarian initiative has constructed a comprehensive database of R-fMRI publications that is maintained in Mendeley (Mendeley, Inc., New York, NY). The database is updated monthly using PubMed-based searches that are run in Sente (Third Street Software, Inc., Denver, CO) with the following query: `‘‘resting state fMRI’’ OR ‘‘intrinsic functional connectivity’’ OR ‘‘rest AND functional connectivity’’ OR ‘‘fmri AND default mode network’’`. Abstracts corresponding to publications returned from the search are hand vetted by CMI Librarian staff to exclude papers that are duplicates, do not have an English language abstract, or are not explicitly related to fMRI and resting state. Based on the content of the abstracts, the articles are then assigned tags based on their topic areas. Articles are marked as *clinical* if they deal with a clinical population of any kind, and are additionally tagged with the specific disorder/population. The *basic neuroscience* tag is applied to papers that focus on non-clinical populations and are additionally categorized as *brain and behavior*, *functional anatomy*, or *multimodal*. The latter refers to papers that integrate fMRI with an additional imaging modality. There are additional tags for papers that incorporate genetics (*genetics*) and those that include animal research (*animal models*). Journal articles that are mainly focused on a particular analytical or imaging technique are classified with the *methodology* tag. Finally there are tags for reviews and meta-analyses (*review/meta-analysis*) and papers that use data from the 1000 Functional Connectomes and International Neuroimaging Data-sharing Initiative (*1000 Functional Connectomes*).

2.2. Publication Trends

Publication rates, venues, subject areas, and open access policies were measured and analyzed using metadata found in the CMI Library and complementary PubMed queries. The growth rates of literature volume over time were found for R-fMRI and compared to that for all of fMRI - the mother discipline of R-fMRI. First, the CMI Librarian was aggregated by year to find each years publication count. Then, the following query was used on PubMed to find the number of articles in all of fMRI per year: `(‘‘fMRI’’ or ‘‘functional magnetic resonance imaging’’) and (‘‘YEAR’’[Publication Date])`. The cumulative sums were then calculated to find the total literature volume over time.

Growth rates were modeled with exponential functions. An exponential growth rate indicates that the literatures growth rate ${}^{dV}\!/_{dt}$ is proportional to its volume (V), ${}^{dV}\!/_{dt} \propto V$. To model publication growth, piecewise exponential functions of the form $V(t) = V_0 e^{Rt}$ were fitted over the intervals Janurary 1, 1994 - December 31, 2005 and Janurary 1, 2006 - December 31, 2012 for both R-fMRI and fMRI. The year 1994 was chosen as the first point because growth is subexponential before this period. Between 1994 and 2012, growth is super-exponential if fitted over the full period; so two piecewise curves were fitted. The break point at 2006 was chosen by minimizing model error over the full period.

Several additional publication trend statistics were computed from CMI Librarian and PubMed metadata. Counts of publications by journal were found directly from CMI Librarian data. Clinical applications were aggregated from the librarys hand-curated tag information. Open access rates for R-fMRI and all of fMRI were found from PubMed.

Using address information from the CMI Librarian, a density-equalizing map, or cartogram, was generated on the publications correspondence addresses with the ScapeToad (http://scapetoad.choros.ch) implementation of the Gastner/Newman diffusion-based algorithm [?]. The goal of the cartogram, and the Gastner/Newman algorithm in particular, is to plot a geographical map such that the density of the publications by country is constant per unit area, and the boundaries of each country are still recognizable.

2.3. Term Frequency Analysis

A bag of words model was used to learn about the fields experimental methods and areas of focus. A bag of words model assumes that all of the words in a corpus are statistically independent, regardless of whether they appear in the same sentence or publication. Under this model, a terms significance can be measured simply by counting its occurrences. **Cameron: Should we reference a text mining review here?**

Terms of interest were derived from n-grams (a phrase consisting of n words) from neuroimaging methods, cognitive ontology [?], and the PubBrain lexicon (http://www.pubbrain.org). N-grams from cognitive ontology and the PubBrain lexicon were found from their respective online sources. A domain expert (RCC) manually generated the list of terms for neuroimaging methods. The final sets of n-grams were found by iteratively testing a candidate set, then combining synonyms and compound terms to form a new candidate set.

For each term, conditional term frequency (conditional-*tf*) was computed as the median ratio of term count to total word count in each publication that contained the term of interest. Document frequency (*df*) for each term was computed as the number of documents containing the term. Terms with high *df* are popular across the corpus, while terms with high conditional-*tf* occur often in the documents in which they appear.

2.4. Prevalence Of Methods

A Naïve Bayes model was used to determine the prevalence of various analysis methods in the literature. **Cameron: Include a citation for this?** In particular, this model was used to discriminate papers that used seed-based correlation, ICA, clustering, graph theory, and machine learning methods. First, a pre-training set of 40 publications was manually created. From these publications, important features were determined by finding tokens and n-grams with high values of the χ^2 statistic, which is used as a proxy for information gain [?]. A term frequency matrix was then constructed, relating each publication to the frequency of each term contained in the publication. Logarithmic term frequencies ($\log tf$) were used to prevent a single term from dominating the model. Each publication's term frequency vector was normalized to unit length to mitigate the effect of publications that contain too many or too few terms. This term frequency matrix was used to train a Naïve Bayes model. The model was evaluated using five-fold cross-validation on 183 additional publications that were also classified by hand.

2.5. Citation Analysis

Publications and the citations that link them were modeled as a directed graph to characterize their relationships and "small world" nature [?]. A directed graph is a general framework used to represent relationships between nodes (publications) using edges (citations), which each point from one node to another.

First, PDF files corresponding to the titles in the CMI Librarian were systematically downloaded. A fuzzy search for every publication title in every file was performed using the SequenceMatcher method of Pythons difflib library (http://docs.python.org/2/library/difflib.html). This fuzzy search first cleaned the binary file with regular expressions to remove junk patterns and normalize whitespace. The longest identical match of the publication title was compared to this match and its context to the true publication title by finding the match ratio (measure of similarity between strings). Of 20,541 total matches with ratio greater than or equal to 0.8, 16,425 (80%) had a ratio of 1.0 i.e., they were exact matches and 17,983 (88%) had a ratio at least 0.9. The cutoff for true matches was set to 0.9, resulting in 17,983 citations.

Using the citations found by the fuzzy search, a graph was constructed of the CMI Librarian publications (nodes) and the citations between them (edges). From this graph, the most central nodes were found using the pagerank implementation in NetworkX [**?**]. Pagerank, a variant of eigenvector centrality, aims to find the steady state probability of a random walk reaching each node [**?**]. These solutions are given by the equation:

$$Pr(p_i) = \frac{1-d}{N} + d \sum_{p_j \in M(p_i)} \frac{Pr(p_i)}{L(p_j)} \qquad (1)$$

where N is number of nodes, d is the probability of continuing the random walk, $M(p)$ is the set of pages that link to node p, and $L(p)$ is the number of outbound edges from p. The solutions, $Pr(p_i)$, are the dominant eigenvector of a modified adjacency matrix, and can be found with the power method.

Finally, a jackknife procedure [**?**] was performed to reduce the variance of the Pageranks and to quantify their sensitivity to choices of fuzzy-search parameters. For each of over 1,000 replicates, 20% of the edges were randomly deleted, and Pagerank was calculated. The Pagerank mean and standard error over the jackknifes were reported, which is an overestimate of the true procedure error. Additionally, the graph's mean shortest path length and mean clustering coefficient was calculated in a seperate analysis.

2.6. Graph Analysis of Co-authorships

Collaborative relationships between authors of R-fMRI literature can be characterized by co-authorships. A graph was constructed from the CMI Librarian data in which each node corresponds to an author and edges represent co-authorships, weighted by the number of papers that the two authors appear on. From this graph, we calculated average path length, clustering coefficient, Pagerank centrality, and the quantity of disjoint sets. Additionally, we tested for graph robustness by repeatedly removing authors and publications and measuring graph connectivity.

2.7. Working Groups

We defined working groups as sets of researchers who frequently publish together and, once found, the publication patterns of these groups were examined. In order to find these groups, a greedy community-detection algorithm was developed that uses the intersection of researchers publication sets to measure their similarity.

1: $WG \leftarrow \varnothing$
2: **for** each author s by decreasing $|\text{pubs}(s)|$ **do**
3:     $a \leftarrow s$
4:     **repeat**
5:         $P \leftarrow P + a$
6:         $a \leftarrow \arg\max_{a \notin P} |\text{pubs}(a) \cap \text{pubs}(P)|$
7:     **until** $|\text{pubs}(a)| < 10$ **or**
8:         $|\text{pubs}(a) \cap \text{pubs}(P)| / |\text{pubs}(P)| < 0.3$
9:     **if** $\sum |\text{pubs}(P_i) \cap \text{pubs}(P)| \geq 50$
10:         **and** $\sum I(|P \cap WG_i| / |WG_i| > 0.2) = 0$
11:             **then** $WG \leftarrow WG + P$
12:     **end if**
13: **end for**

The procedure for identifying author working groups begins by repeatedly seeding a new potential working group P with the seed author s who has the most publications (lines 2-5). Authors are searched to find the author a with the most co-authored publications with the working group (line 6). If this author has more than 10 publications, and is a co-author on at least 30% of the publications authored by the members of the working group (line 7), then the author a is added to the potential working group P (line 5). Otherwise, the potential working group is closed and considered for inclusion in the results. If the sum of all papers in common across the potential working group is greater than 50 (line 9), and the potential working group has fewer than 20% of authors in common with any other working group (line 10), then it is included in the results (line 11).

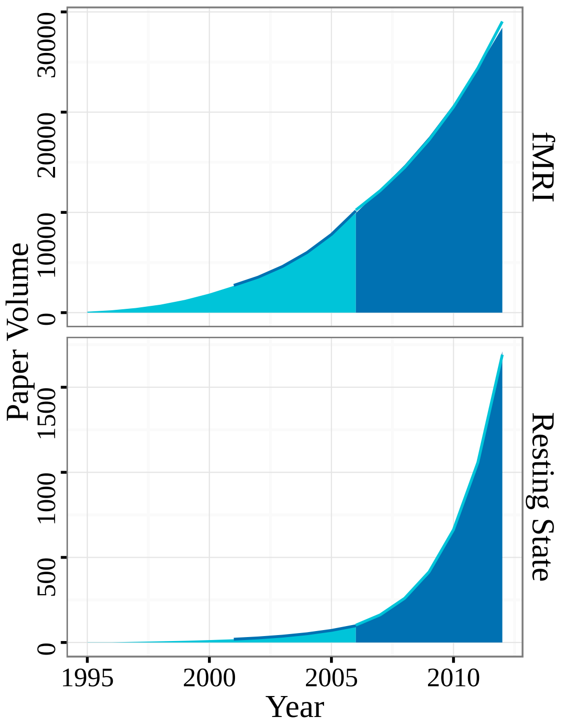

Figure 1: The growth of resting state and fMRI literature.

3. Results

3.1. Publication Trends

Growth of R-fMRI literature was fitted by piecewise exponential functions with 32% growth between January 1,1994 and December 31, 2005 and 47% between January 1, 2006 and December 31, 2012. In comparison, growth of fMRI literature was found to be 26% and 17% during the same time periods. Fig. **??** shows the growth of each literature, and represents a total of 28,434 fMRI publications and 1,721 R-fMRI publications. Additionally, 39% of R-fMRI publications were determined to be open access, compared to 31% in all of fMRI.

The top 20 publication outlets for R-fMRI publications accounted for 59% (1017) of the CMI R-fMRI library as shown in Fig. **??**. This list is dominated by three neuroimaging journals (NeuroImage, Human Brain Mapping, Brain Connectivity), which accounted for approximately 26% of the R-fMRI publications. The top 20 also include nine general neuroscience, two general science, four clinical neuroscience, and two general imaging journals which accounted for approximately 16%, 10%, 5%, and 3% of the library respectively. The remaining 41% of the library was spread over 268 different publication outlets. Nine of the top 20 journals began publishing R-fMRI literature before 2006, and Brain Connectivity, founded in 2011, is the most recent addition.

The density-equalizing map is shown in Fig. **??**. The top three countries are the United States of America, United Kingdom, and Netherlands, wherein lie 62.9%, 9.6%, and 8.6% of correspondence addresses, respectively. **Cameron: Re-color the density map so that the colors are meaningful, and are more attractive.**

3.2. Term Frequency Analysis

Growth of the most prevalent clinical terms is shown in Fig. **??**. Of 1,721 publications in the corpus, "Clinical" is the most commonly applied tag (Tab. **??**). "Schizophrenia" (13%), "Alzheimer's Disease" (11.3%), and "Depression" (10.8%) are the most common clinical sub-tags that co-occur with the "clinical" tag.

Use of the term "connectome" has steadily increased across several measures since 2009 . Term frequency (*tf*) has increased at a rate of 137 mentions per year, document frequency (*df*) at 30 documents per year, and term frequency-inverse document frequency ($tf \times df^{-1}$) at 33 terms per document per year.

Fig. **??** shows conditional-*tf* and *df* for the terms with the highest values. The most common imaging modality was fMRI, which was a key term in the searches used to generate the corpus. The most investigated cognitive domains were activation, memory, attention, and association. The PFC, PCC, and anterior cingulate were the most discussed brain regions.

3.3. Methods Analysis

The accuracy of the Naïve Bayes model in inferring algorithmic methods is shown in Tab. **??**.

The terms used as features, along with their χ^2 values, are shown in Tab. **??**.

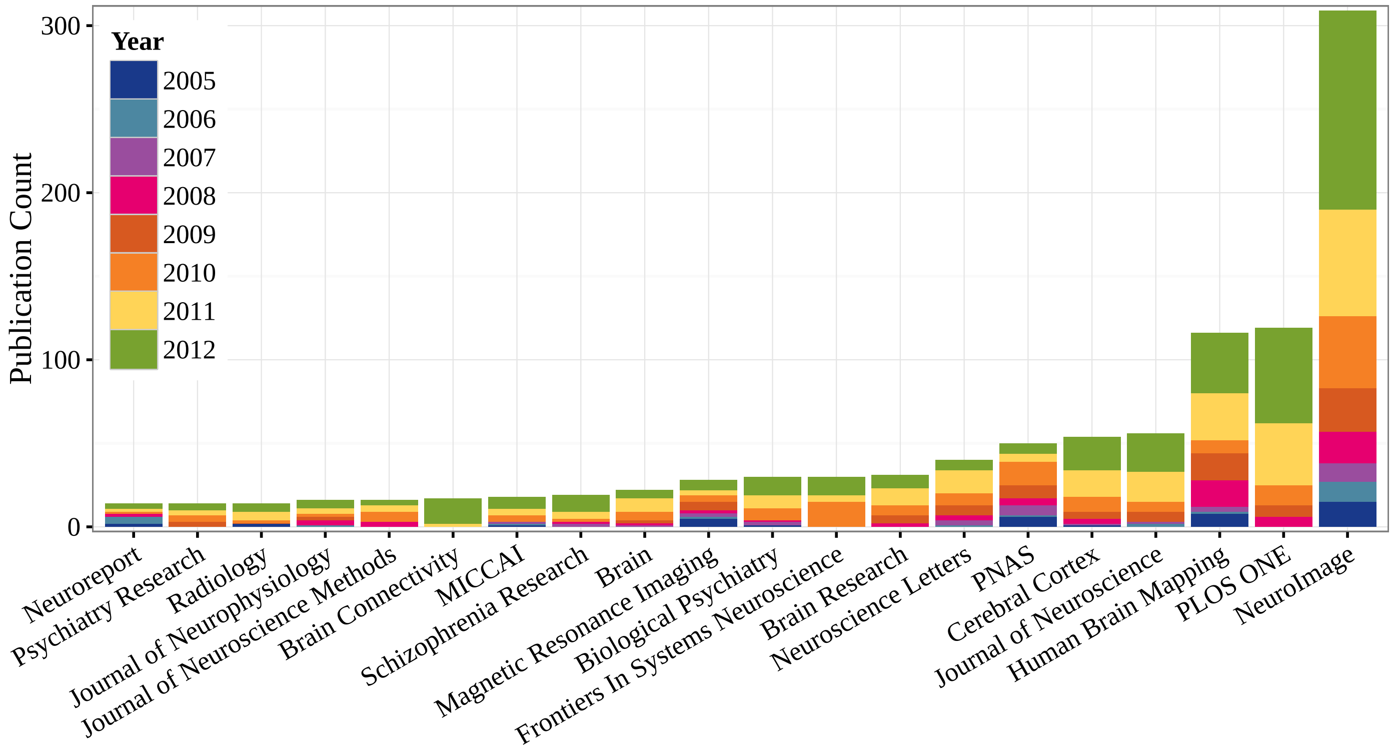

Figure 2: Publication counts by journal.

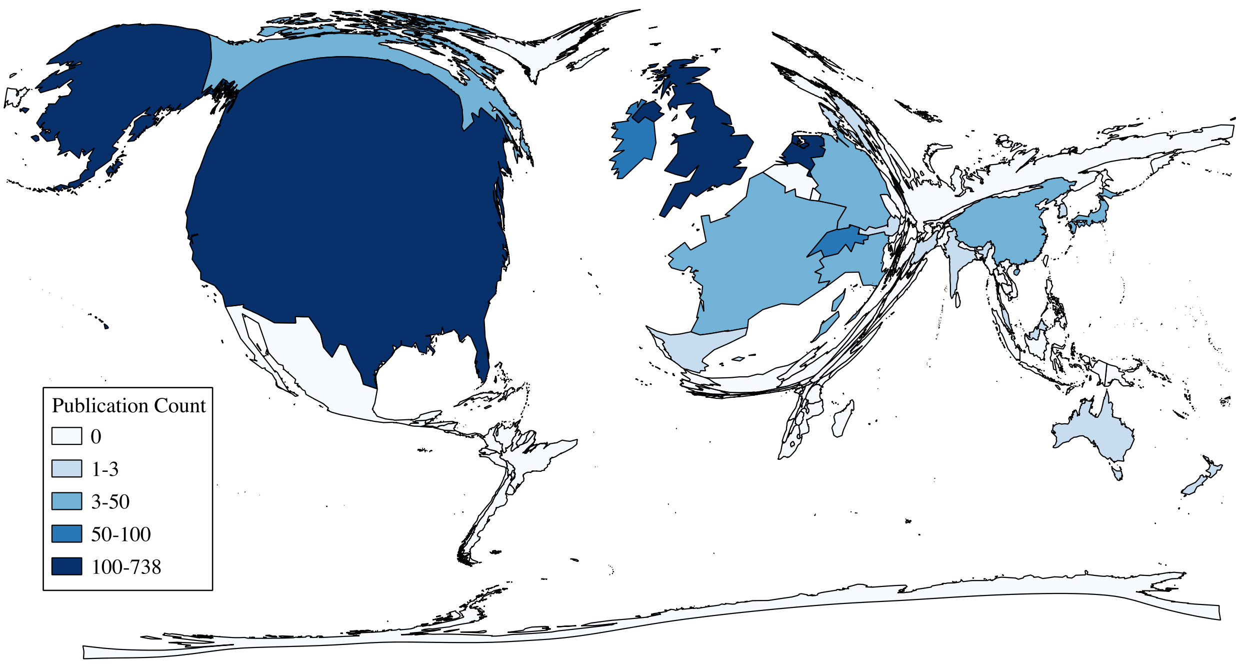

Figure 3: Number of resting state publications by country.

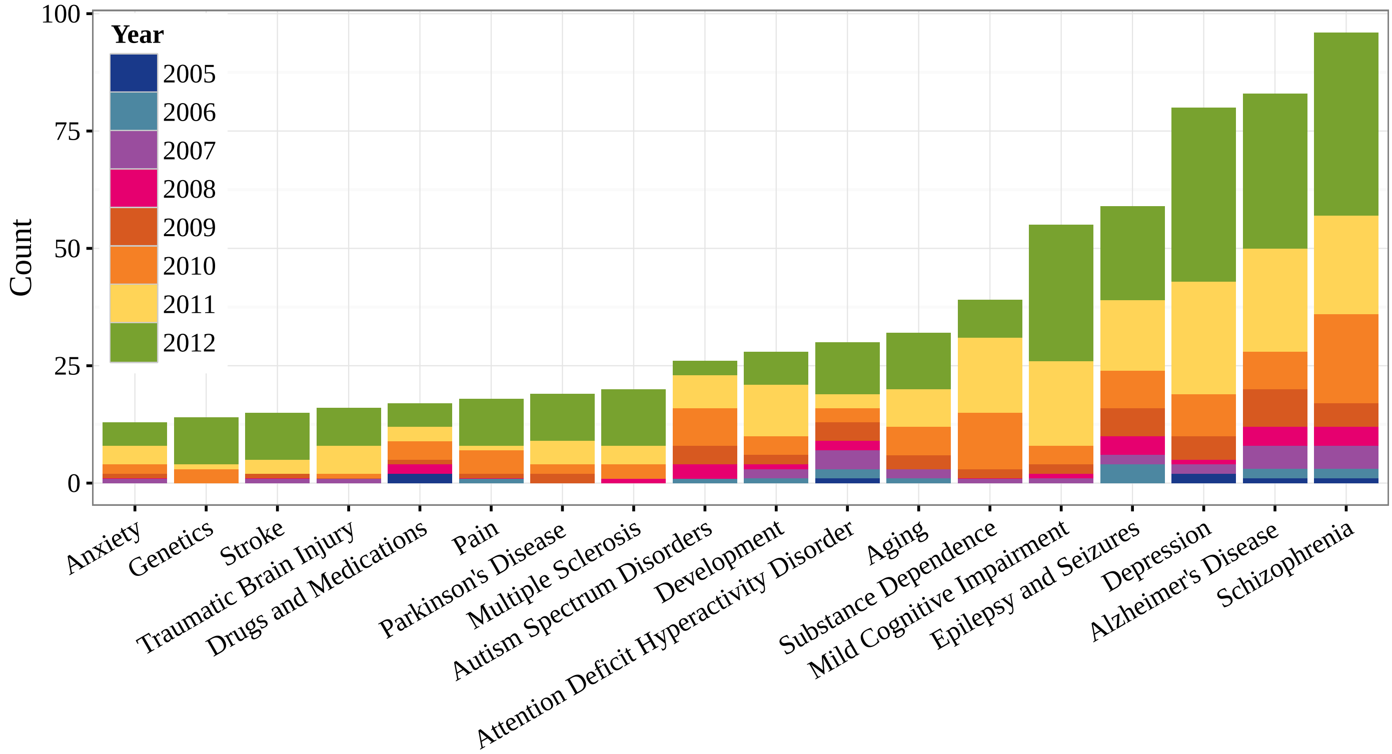

Figure 4: Most common clinical terms from the resting state literature.

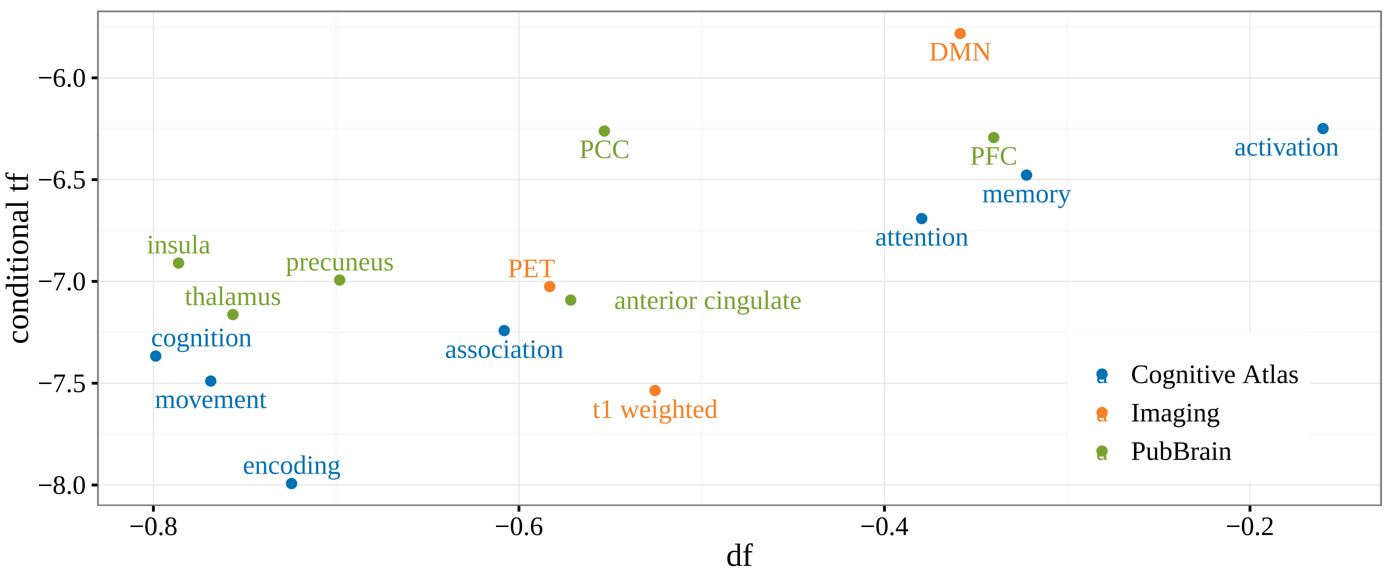

Figure 5: DF and conditional-TF for terms with the greatest values of each

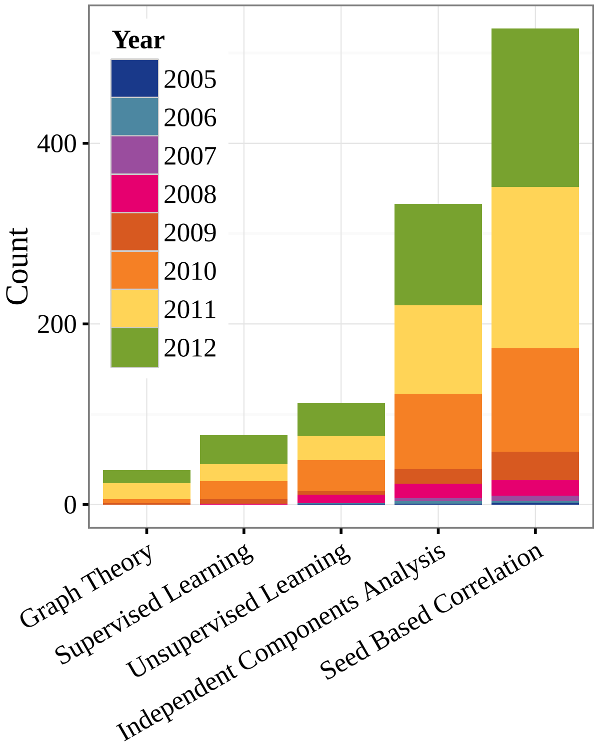

Figure 6: Prevalence and growth of different methods applied in the resting state fMRI literature.

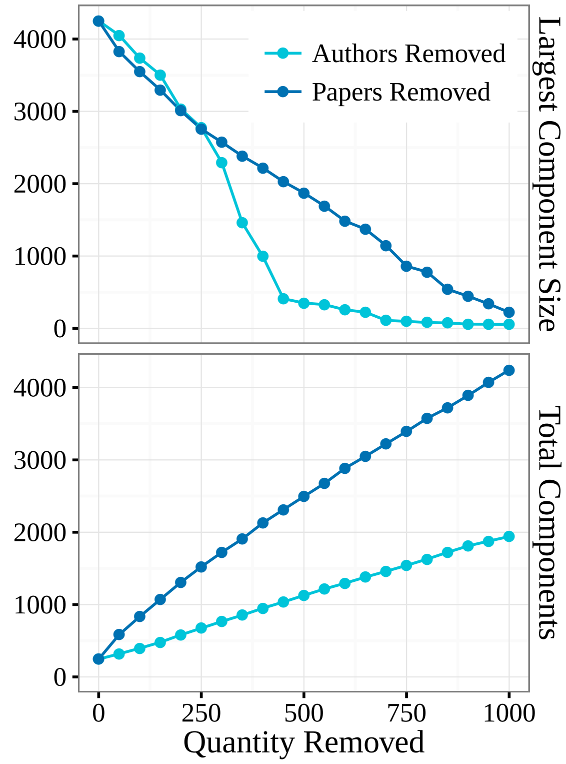

Figure 7: Number of components and size of largest component after removing authors and papers from co-authorship graph.

Fig. **??** shows growth over time of algorithmic methods in R-fMRI literature. Seed-based correlation is by far the most common method, accounting for nearly half of all uses of algorithmic methods. Independent component analysis (ICA) and clustering have shown steady growth since 2010, and use of machine learning methods increased at a faster pace in 2012.

3.4. Citation Analysis

Cameron: I think that the citation analysis works better as a table (??) than a graph. We have space to add the next 10 papers, and the citation counts. also add in gephi plot of citation network

The top 10 publications by pagerank were collectively cited by 66% of the corpus, and the top 1% of publications account for 10% of the total pagerank. After these publications, pagerank falls off more slowly, with the next 10% of publications accounting for 40% of the total pagerank.

The mean clustering coefficient was 0.094 (standard error 0.010) and the mean shortest path length was 4.4 (standard error 0.080). In a set of 1,000 random graphs constructed on the same set of vertices, the mean clustering coefficient was 0.014 (standard error 0.0030) and the mean shortest path length was 4.1 (standard error (0.60).

3.5. Co-authorship Analysis

The co-authorship graph was found to be robust, exhibiting significant small-world statistics. In particular, its clustering coefficient was 0.878, compared to 0.002 for a random graph with the same number of nodes and edges, and its average shortest path length was 4.964, compared to 3.915 for the random graph.

Fig. **??** shows the number of connected components, and size of the largest connected component, as publications and authors are removed.

3.6. Working Groups Analysis

Four working groups were identified. Their combined 23 authors (0.6% of 3,704 across the

Table 4: Publication Pageranks

#	Title	Year	Pagerank (SD)	Citations
1	A default mode of brain function.[?]	2001	0.0165 ($3\cdot10^{-4}$)	MISSING
2	Functional connectivity in the motor cortex of resting human brain using echo-planar MRI.[?]	1995	0.0136 ($3\cdot10^{-4}$)	MISSING
3	Functional connectivity in the resting brain: a network analysis of the default mode hypothesis.[?]	2003	0.0104 ($2\cdot10^{-4}$)	MISSING
4	Consistent resting-state networks across healthy subjects.[?]	2006	0.0098 ($2\cdot10^{-4}$)	MISSING
5	Investigations into resting-state connectivity using independent component analysis.[?]	2005	0.0069 ($2\cdot10^{-4}$)	MISSING
6	Functional connectivity in single and multislice echoplanar imaging using resting-state fluctuations.[?]	1998	0.0056 ($2\cdot10^{-4}$)	MISSING
7	Default-mode network activity distinguishes Alzheimer's disease from healthy aging: evidence from functional MRI.[?]	2004	0.0056 ($2\cdot10^{-4}$)	MISSING
8	Toward discovery science of human brain function.[?]	2010	0.0055 ($2\cdot10^{-4}$)	MISSING
9	Dissociable intrinsic connectivity networks for salience processing and executive control.[?]	2007	0.0055 (10^{-4})	MISSING
10	Intrinsic functional architecture in the anaesthetized monkey brain.[?]	2007	0.0047 (10^{-4})	MISSING

literature) cover 17.5% of R-fMRI publications. The mean publication count by authors in these groups was 21.9; overall, it was 0.31. The mean number of publications in common between pairs of authors in the same working group was 10.7; across different working groups, it was 0.11.

Working group authors (number of publications in corpus by author):

1. Michael P. Milham (41), F. Xavier Castellanos (37), Bharat Biswal (36), and Clare Kelly (36)
2. Tianzi Jiang (35), Kun-Cheng Li (27), Chunshui Yu (19), Li-Xia Tian (18), Yong Liu (15), and Yuan Zhou (15)
3. Qi-Yong Gong (35), Yi-Jun Liu (21), Hua-fu Chen (20), Wei Liao (20), Guang-Ming Lu (18), Zhiqiang Zhang (17), and Yuan Zhong (13)
4. Bradley L. Schlaggar (19), Steven E. Petersen (16), Alexander L. Cohen (13), Damien Fair (13), Nico U. F. Dosenbach (10) and Fran M. Miezin (10)

4. Discussion

Our bibliometric analysis of R-fMRI literature lends insight into the current state of the field, demonstrating its strength, areas of focus, and future potential. The growth of R-fMRI literature is currently significantly faster than fMRI, though R-fMRI currently comprises a small fraction of the total volume of fMRI literature. A high rate of publication growth likely indicates not only that researchers in the field are publishing more, but also that the field is growing to include new researchers. The journal Neuroimage has published more R-fMRI literature than any other. In general, nearly 50% of the literature was published in general neuroimaging journals, perhaps reflecting the rapid advancement of methods used to study R-fMRI over the past 10 years.

Cameron: Need a table with these counts.

In contrast, only 1% was published in clinical journals, suggesting that the field has not yet matured enough to be used in clinical applications. But despite their venues, a full 35% of the R-fMRI corpus was tagged Clinical. Indeed, a great proportion of R-fMRI research is performed on both clinical and control populations in order to identify differences between them; typically, however, the results of these experiments are not immediately applicable in clinical medicine. **Cameron: this is a odd statement**

Another pattern that was explored, open access, is a relatively recent innovation that allows researchers to access publications without payment to the journal by either front-loading the cost onto the author or subsidizing the cost in another way. Open access is not universal in R-fMRI (26%) or fMRI (22%), but has strong footholds in both. That the rate of open access is consistently increasing over time in both literatures bodes well for the future of open science. **Cameron: May want to include information about the NIH mandate that publications resulting from their grants be published open-access.**

Word frequency analysis easily identified the corpus domain, fMRI, as the most common imaging modality. In addition, there was a focus on the prefrontal cortex (PFC), which is implicated in executive function, as well as on the posterior cingulate cortex (PCC), which is a central node of the DMN **Cameron: citation here**. Notably, neither the parietal cortex nor any of its subregions appear in the top hits, though they are also implicated in the DMN. This result may demonstrate the bag of words models limited ability to aggregate concepts: for example, it may be the case that the sum of occurrences of subregions of the parietal cortex would appear in the top hits, but none appear individually. Finally, memory and activation were the most discussed cognitive domains, reflecting current research trends such as in **Cameron: missing citation [8]** and **Cameron: missing citation [9]**. A possible extension of this research is to identify common terms across scientific fields to determine, for example, whether particular genes can affect PFC and PCC function **Cameron: this is not clear**.

Analysis of working groups found particularly prolific labs and collaborators, offers measures for whether the field is dominated by only a few authors, and found patterns in geographic localities and language. The resulting groups are together responsible for nearly a fifth of the total corpus, including some of the highest-Pageranked publications, though there is little publication overlap between groups. Ranking authors by their impact factor reveals that several of the authors in these groups, including M. P. Milham, F. X. Castellanos, C. Kelly, B. Biswal, and Qi-Yong Gong, are not only prolific, but also are attributed many citations. Furthermore, geographic location, language, and affiliation are consistent within groups, but very different between groups: M. P. Milhams group is based in New York, T. Z. Jiangs group is based in Beijing, Qi-Yong Gongs group is based in Sichuan, and Bradley Schlaggars group is based in St. Louis. This result demonstrates the strength of R-fMRI research around the world, and suggests that distance, time zone, and language pose practical barriers to collaboration **Cameron: this statement is pushing it**.

Analysis of the citation graph concluded that it can be considered "small world" because its mean clustering coefficient is significantly higher than that of a random graph, and its mean shortest path length is not significantly different from that of a random graph [**?**]. In a small world network, most nodes are not adjacent to one another, but the mean path length is still fairly short because there are few degrees of separation needed to reach one node from another. More precisely, the mean path length of a small world network grows with the logarithm of the total number of nodes. In this application, this result indicates that although most publications do not cite a large fraction of the corpus, but the number of hops needed to traverse from one publication to another is still fairly short. **Cameron: What qualifies as a large fraction here? We should offer something better to back up this statement** Indeed, there are many real-world relationships, such as social networks and Internet connections, that can be represented with small world graphs.

By measuring Pagerank, publications were identified that are central to R-fMRI literature without necessarily having the highest raw citation count. Among the highest Pageranked publications are the seminal publications that provide much of the historical groundwork for the field. Indeed, from these publication emerges a narrative of R-fMRI history, from the foundations of R-fMRI functional connectivity and activation, to the first clinical application in the field, to the discovery of the canonical independent R-fMRI networks. Relatively consistent Pageranks among the remaining publications suggest that the field is not completely represented in only a small subset of the corpus, but rather is still growing in high-impact directions.

The publication with the second highest Pagerank, Biswal et als "Functional connectivity in the motor cortex of resting human brain using echo-planar MRI" (1995)[**?**], demonstrated functional connectivity in spontaneous, low frequency activity measured using R-fMRI. Resting state functional connectivity had been observed in clinical settings for a long time using electroencephlogram (EEG), and positron emission tomography (PET), which measures glucose metabolism. However, this was the first time it was observed in the complex fMRI-measured BOLD signal, which lacks clear physiological underpinning. Bharat Biswal appears in a top working group and in the list of highest impact authors. Functional connectivity, first demonstrated with R-fMRI in this publication, is among the most frequently occurring terms in the corpus **Cameron: the corpus is determined by the presence of the term functional connectivity**.

Another, separate line of research took root with a series of 2001 publications that include those with first and seventh highest pagerank, Raichle et als "A default mode of brain function," and Gusnard et als "Searching for a baseline: functional imaging and the resting human brain." These publications established that decreases of brain activity that occur during rest are spatially consistent, and introduced the notion of a functional default mode. This line of research on activation was cognitive neurosciences contribution to the foundations of R-fMRI theory. Indeed, activation was found to be among the most frequent cognitive terms in the corpus.

These disparate contributions from neuroimaging and cognitive science were married in the third highest pageranked publication, Grecius et als "Functional connectivity in the resting brain: a network analysis of the default mode hypothesis" (2003). Grecius showed that, in fact, the regions that show correlation during rest are also activated during rest. Thus, this publication introduced the notion of the default mode network, whose regions show both low frequency correlation of spontaneous activity during rest, and also decreased task-related activity. Notably, the default mode network is among the most common neuroimaging terms in the corpus.

In 2004, Grecius followed with the fifth highest pageranked publication, "Default-mode network activity distinguishes Alzheimers disease from healthy aging: evidence from functional MRI." This work linked clinical neuroscience with R-fMRI functional connectivity by demonstrating with fMRI not only that resting state activity was reduced in Alzheimers patients, but also that the regions in which R-fMRI activity was reduced were the very same that had been previously implicated in Alzheimers disease.

In 2006, DeLuca et als "fMRI resting state networks define distinct modes of long-distance interactions in the human brain" found resting state networksregions that are individually correlated during restthat are independent both functionally and statistically. These networks have been applied widely since their discovery, and appear in the list of most common neuroimaging terms in the corpus.

5. Conclusion

Our bibliometric analysis demonstrates that, though still small compared to fMRI, R-fMRI is a rapidly-growing field with major international research hubs. The research community is tight-knit, as shown by its small world citation network, and the field is no one-trick pony, as shown by the large number of highly-pageranked publications. The word frequency analysis identified key concepts in use, and suggests future paths toward integration of R-fMRI with other rapidly-growing fields such as genetics.

6. Acknowledgements

This research was supported by the Child Mind Institute Endeavor Scientist program.

References

Table 1: The number and fraction of publications in the library tagged with a specific keyword

CMI Librarian Tag	**Frequency**
Clinical	747 (43%)
Basic Neuroscience	593 (34%)
Meta-analysis/reviews	235 (14%)
Methodology	210 (12%)
Multimodal	143 (8%)
Brain and Behavior	79 (5%)
1000 Functional Connectomes	61 (4%)
Animal Models	61 (4%)
Functional Anatomy	44 (3%)
Genetics	29 (2%)

Table 2: Mean cross-validation accuracy of the Naïve Bayes model

Accuracy	89.8%
Sensitivity	85.8%
Specificity	90.9%
True Positives	169.9
True Negatives	671.8
False Positives	67.1
False Negatives	28.1

Table 3: χ^2 values for the top features used by the model.

Top Features	
Seed-based Correlation	χ^2
seed	148
seed region	90
seed-based correlation	49
ICA	χ^2
ICA	198
independent component	112
cross-validation	182
RSN	80
Clustering	χ^2
clustering	86
fractal	68
k-means	65
exponential	44
Graph Theory	χ^2
graph	144
small world	118
edges	112
topological	100
modularity	96
worldness	84
graph theory	82
undirected	81
hubs	62
centrality	60
Machine Learning	χ^2
Vapnik	151
sigmoid	112
multiclass	112
machine learning	92
support vector	78