

# Examining a Quarter Century of Publishing Trends in Social Work Research: A Data Science Perspective

May 26, 2015

## Data Revolution

It is estimated that 90% of all the data in human history has been generated in the past two years, and the amount of data continues to grow by roughly 2.5 quintillion bytes every day (IBM, nd-a). We have unprecedented opportunities for deriving actionable insights from these data to advance the human condition. The significance continues to gain the attention of national and international organizations. For example, the Independent Expert Advisory Group (IEAG) of the United Nations prepared a report entitled, *A World That Counts: Mobilising the Data Revolution for Sustainable Development* (IEAG, 2014). The IEAG (2014) argues that “Data are the lifeblood of decision-making and the raw material for accountability. Without high-quality data providing the right information on the right things at the right time; designing, monitoring and evaluating effective policies becomes almost impossible” (p. 2). The United States federal government established an open data initiative, motivated by the awareness that data can have the greatest impact when it is “accessible, discoverable, and usable” ([www.data.gov](http://www.data.gov)). Data from all levels of government that is made completely open can lead to “cost savings, efficiency, fuel for business, improved civic services, informed policy, performance planning, research and scientific discoveries, transparency and accountability, and increased public participation in the democratic dialogue” ([www.data.gov](http://www.data.gov)). At the time of preparing this article, over 124,000 data sets were freely available through [www.data.gov](http://www.data.gov) among a diverse set of topic areas: agriculture, business, climate, consumers, ecosystems, education, energy, finance, health, local government, manufacturing, ocean studies, public safety, and science and research.

In 2012, the NIH established the Big Data to Knowledge (BD2K) initiative, recognizing the potential of data to advance our understanding of human health and disease. This initiative was designed to support four broad goals: 1) to facilitate use of biomedical digital assets by making them discoverable, accessible, and citable; 2) to conduct research and develop methods, software and tools to analyze biomedical big data; 3) to enhance training in the development of use and tools necessary for biomedical big data science; and 4) to support a data ecosystem that accelerates discovery as part of the digital enterprise (NIH). At roughly the same time, the National Science Foundation (NSF) created new initiatives to harness the potential of the data revolution, arguing that “data are motivating a profound transformation in the culture and conduct of scientific research in every field of science and engineering . . . American scientists must rise to the challenges and seize the opportunities afforded by this new, data-driven revolution” (Suresh, XXX). The programs of the NSF are organized around capacity building and innovative applications. Capacity building activities focus on developing fundamental theories, techniques, methodologies and technologies that are broadly applicable to big data problems. Innovative applications relate to specific domains advancing innovative ideas that provide solutions “with potential for a broader impact on data science and its applications.”

## Data Science and Social Work

We are indeed living in a data revolution. The relevance to social work research is clear — we have the opportunity to study human behavior and other social phenomena in real time, with vastly improved reliability in our measurements, and at scale. Data comes from a wide range of sources, including (but not limited to) social media interactions, electronic medical records, audio and video recordings, click stream data, transaction data, web logs, embedded sensors, and GPS. New tools allow blending of data sets, unique ways to analyze data, and methods of visualizations that can produce insights limited only by one’s imagination. As stated by the NIH, “lack of appropriate tools, poor data accessibility, and insufficient training, are major impediments to rapid translational impact” in this new era of data. It is essential that the field of social work fully embraces the data revolution and considers the extent to which the existing research infrastructure and

training is preparing both social work researchers and providers to be effective and efficient in a data rich world. The potential of the data revolution is reflected in the current *Grand Challenge* entitled, “Harness technology for social good.” However, we believe that the data revolution goes beyond this single *Grand Challenge* as it provides the opportunity to be informative in all areas of social work research.

Data science serves as a useful framework for broadening the existing research infrastructure and training opportunities within social work. Data science is similar to traditional academic research, as it includes the key roles of substantive expertise and traditional research methodologies. Much of the theoretical basis comes from the field of statistics (Zumel & Mount, 2014), but the famous statistician William Cleveland argued that it is an interdisciplinary field that is much larger than statistics (Cleveland, XXX). A major difference is that data science draws heavily from knowledge in the areas of computer science, software engineering, and information technology. Unlike traditional academic research, data science is particularly well suited for tackling the 3 V’s of big data: *volume*, *variety*, and *velocity*. More specifically, in data-rich environments, researchers encounter an amount or volume of data that often exceeds the capacity of traditional database systems. Data also come in a variety of formats, ranging from highly structured data contained in relational databases to massive stores of unstructured text-data culled from social media. Many data opportunities go beyond the traditional approach of downloading a dataset and analyzing it with an off-the-shelf statistical package like SPSS, Stata, or SAS. Now, data can be accessed in real time through application program interfaces (API), which brings along the challenge of *velocity* — that is, data can be transmitted at a rate that grossly exceeds the computing power of desktop computers. These are practical challenges that need to be addressed in order to maximize the potential of these new data.

Another feature that distinguishes data science from traditional academic research is the focus on generating *actionable insights* from data rather than theory building and hypothesis testing. In fact, data science is more interested in deriving useful insights from correlational patterns as opposed to uncovering causal mechanisms. The focus is on the development of data products (Davenport, XXX), as opposed to reporting analysis. Data products are tools or services that generate actionable insights from the data itself, including (but not limited to) automated reports or dashboards, recommendation systems, prediction algorithms, decision tools, and interactive visualizations that help people with non-technical background make sense of complex data. Data science also values the philosophy of *openness* or open science. Thus, much effort is devoted to creating sustainable systems that make data as widely and freely accessible as possible, and developing and promoting the use of open source tools for managing and analyzing data. The concept of reproducible research is a *rule* rather than the *exception*. This involves making both data and statistical code available alongside research reports. This helps ensure the quality of the research, in addition to allowing researchers to make advances without having duplicate work that has already been done. Data science is by no means incompatible with traditional academic research. We believe that the tools, research strategies, and principles of data science are complementary to traditional ways of knowledge building in social work, offering many unique opportunities to improve the quality, efficiency, and impact of this area of research.

## Overview of Current Study

In the current study, we seek to utilize the opportunities of the data revolution and data science by applying a range of tools and strategies to construct an historical base of social work research. In doing so we take on two of the 3 Vs of big data: volume due the amount of data collected and variety in terms of the multiple sources of that data. The data are article records harvested from existing bibliographic databases hosted on ProQuest and EbscoHost. Article records are already widely used in various bibliometric studies, particularly those that are intended to reveal publication networks within scientific communities. It does so by creating links between co-authors, and then visually representing the edges and vertices of the network. This is also done with article citation histories, to show influence over time. In this study, we were interested in the various article meta-data that are common to many scientific article records. For example, a single article record in psycINFO can have as many as XX pieces of meta-data including: dates, journal titles, article titles, author names, author affiliations, full text abstracts, keywords, subject classification, methodology, location, populations, page numbers, and source of funding. This is a rather large amount of information about a single paper, at least compared to the amount of information contained in article records before

the availability of electronic storage and access. Large collections of article records may contain potentially interesting trends in the research, assuming the selection of the article records were conducted in a way that allowed meaningful inferences to be derived.

In the current study, we sought to capture every journal article record published in every social work journals over the in past quarter century (1989-2013) and to explore the various meta-data to uncover potentially interesting trends in the social work scholarship. For example, we can reasonably infer estimates of the *size* of social work scholarship by counting the number of unique article records. And, growth of the field can be inferred by examining the number of article records over time. Similarly, we examine the extent to which *team science* has become part of the research practice, as this has implications for training of social work researchers. We also use location meta-data to understand the extent to which different areas of the world have and have not been the focus of social work research. Finally, we used topic modeling to extract topic areas that define the focus of social work research. The execution of this research was grounded in the tools and values of data science. Thus, all data were managed and analyzed using the open source software, R. We also used an open source authoring system that integrates the statistical analysis with the text of the manuscript. Along with this manuscript, the authors will also release the actual data and statistical code, thereby meeting the requirements of *reproducible research*, as defined by King (). To our knowledge, this is the first time fully reproducible research (as defined by King, XXXX), has been published in a social work journal.

Davenport: <http://blogs.wsj.com/cio/2014/06/25/so-you-want-to-build-a-data-product/>  
<http://www.nsf.gov/pubs/2015/nsf15544/nsf15544.htm>  
[http://www.nsf.gov/news/news\\_summ.jsp?cntn\\_id=123607&org=NSF&from=news](http://www.nsf.gov/news/news_summ.jsp?cntn_id=123607&org=NSF&from=news)

Why is this study important:

1. Substantive contributions Understand publication trends from a macro perspective. This can:

From an economic standpoint, the field of social work observed growth throughout the economic crisis and is expected to maintain growth. The number of MSW programs also continues to grow. While growth is observed in a variety of areas, it is important to consider the extent to which growth has occurred regarding overall scholarship.

From a methodological standpoint, this study makes a few unique and important contributions to the social work research. Foremost, the study was designed and executed in a manner that meets the full standards of reproducible research. The CTSPEDIA, an advocate of fully reproducible research, defines this type of research as, "the practice of distributing, along with a research publication, all data, software source code, and tools required to reproduce the results discussed in the publication. As such the RR [reproducible research] package not only describes the research and its results, but becomes a complete laboratory in which the research can be reproduced and extended." (NOTE ABOUT ORGMODE)

Methods

Include section on reproducible research.

Discussion

## Growth of Social Work Research

From the years 1989 to 2013, the field of social work research produced nearly 35,000 journal articles. This doesn't include books or book chapters, editorials, book reviews, and other types scientific communications. The growth increased steadily from XX to XX, and appears to be leveling off. Through this period of time, we observed XX unique titles appear while a number of other journals were lost.

The current study helps clarify the extent to which *team science* is emerging in social work research. This is an important issue to consider, as it has implications for training researchers, reviewing and understanding tenure and promotion cases, and research infrastructure to facilitate collaborations ([http:](http://)

<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3321548/>). According to Thomson Reuters, the average number of authors on papers included in the Science Citation Index increased by about 50% between 1990 and 2010 ([http://www.hopkinsmedicine.org/institute\\_basic\\_biomedical\\_sciences/news\\_events/articles\\_and\\_stories/employment/2013\\_01\\_middle\\_authors.html](http://www.hopkinsmedicine.org/institute_basic_biomedical_sciences/news_events/articles_and_stories/employment/2013_01_middle_authors.html)) In our research, at least one-half of all published studies were authored by a single individual until 2009. The standard deviation continues to increase, providing further evidence of continued team science growth in social work. How authorship is assessed is a very important consideration for the field, particularly with respect to tenure reviews. It is not clear how reviewers of tenure cases are viewing team science contributions and *middle authors*. However, this is something that will have to be considered for the field. In the field of physics, the weighted average seniority and average seniority measures both had a negative relationship with the impact, based on the number of citations the publication received, as well as a negative relationship between the number of authors and the number of citations. <http://myweb.fsu.edu/bstvilla/papers/AuthorTeamDiversityImpact.pdf> (\*\*NOTE TO BRYAN – AS A QUICK STUDY, YOU AND I SHOULD BUST OUT A STUDY INVOLVING THIS. WE CAN DO IT FOR JUST THE TOP JOURNALS IN SOCIAL WORK THAT ARE IN WOS.)

<http://dmm.biologists.org/content/7/3/311.full>

[https://ccrod.cancer.gov/confluence/download/attachments/47284665/teamscience\\_fieldguide.pdf](https://ccrod.cancer.gov/confluence/download/attachments/47284665/teamscience_fieldguide.pdf)

One thing to consider is the extent to which social work research is being served by other journal outlets. For example, social work researchers with primary research interests in areas of health, mental health, and substance use may publish in speciality journals outside of social work for a variety of reasons. Thus, the amount that social work has grown should be cautious, keeping in mind this particular metric refers only to social work journals. The field may have grown more rapidly in other areas that are not reflected in the current research. The functions that were constructed to *wrangle* the abstracts are functional for records extracted from PubMed. Since this data source contains author affiliations, it would be feasible to extract this information using a set of regular expressions.

According to the journal *Nature*, from the late 1600s until about 1920, the rule was one author per paper. The rule was “breached in the 1920s, diminished in the 1950s, and largely abandoned by the 1980s” (p. <http://www.nature.com/nature/history/full/nature06243.html>). Social work research had a much different trajectory. In social work research, the rule wasn’t breached until 2009.

This has implications for tenure, and it is unclear how *team science* is actually being considered in context of these reviews. The field is arguably in a time of transition, from si

Defining a *disciplinary* social work journal was a complicated task. Social work is inherently inter-disciplinary. We based our definition of a social work journal using the list of Hodge and Lacasse (2011), who created a comprehensive list of journal titles. This list was based on earlier attempts to identify all the disciplinary social work journals. We supplemented our list by doing supplemental searches in all data sources for journals that used any of the following terms.

The length of the articles was examined, as it has implications for device formatting. Clearly, the field of social work has a current average of 15 typeset pages. This consistency allows for targeted customization for mobile devices. An article can be optimized and fairly standardized, which has important benefits for information retrieval and comprehension.

This article raised many opportunities that became obvious in this study. In the current study, we worked in earnest to present as much of the analyses visually. At the time of writing this article, the majority of social work journals do not have full color articles. Indeed, this is a residual policy from when journals were exclusively in black-and-white and before electronic archiving. Now, the majority of research is disseminated electronically, and even the most primitive computers connected to the Internet have color monitors.

Data quality – is there anything to talk about on this topic???

Team science – is it good for social work research?

## Reproducible research

Reproducible research is essential to ensuring the integrity of the scientific record. Making research reproducible is a complicated process that does carry risks. More specifically, it does involve taking extra time to prepare documentation, organize code, and prepare files to be made publicly available. This problem has already been given careful attention in other fields, which can be considered reasonable solutions for solution work research. For example, this study utilized Rmarkdown as the tool for producing reproducible research. Rmarkdown is ... [provide *short* description here – tell reader minimum she/he needs to know].

Another practical consideration of reproducible research is curating the data, and making it available. The availability can be challenging. For example, many social work studies involve sensitive data that ultimately requires very strict limitations on how and where the data are stored, along with restrictions on who can access the data. [Describe the PS journal that helps address this issue ...] This study involves proprietary data from EbscoHost and ProQuest, which prevents us from making all the data available. However, we provide a minimum random extract of the data to ensure verification of our computer code. Additionally, we have made all aspects of our computer code available that will allow researchers with access to EbscoHost and ProQuest to replicate our findings, as well as use the code to build upon this research.

We believe that social work journals should consider implementing policies that encourage reproducible research. Making this best practice in publishing represents a straightforward way to improve the quality of social work research. Along with a policy of reproducible research, the issue of how the journal responds to errors will also be necessary. We certainly do not advocate that authors are expected to retract articles or face public shame in the context of random human error. Some studies may involve generating thousands of lines of computer code – and, a simple bug in the code can produce erroneous results. Researchers who adhere to the principles of reproducible research are ultimately working in earnest to ensure they are producing the highest quality research possible. Errors that are identified in the process represent an opportunity to further advance knowledge, which sometimes means correcting knowledge. These situations simply reveal that the scientific method is actually working the way it is supposed to. At present, we have to assume that all the social work research published to date is without error. In absence of reproducible research, we can not fully assess the quality of research or replicate a study.

Discussion point:

Another complication is the possible consequences that a researcher might experience when an error is discovered. This is an important issue that the field needs to address.

very research report is subject to some form of human error. It is important to consider some of the different types of error and how the field should respond.

(Conduct and analysis on reproducible research" and "replication" or "replicate")

(original citation from orgmode.org: <http://orgmode.org/worg/org-contrib/babel/intro.html>)

conforms to all aspects of what is regarded as reproducible research by making available all the code produced in order to conduct the study.

. Methodological contributions

## Methods

### Extraction of article records

Our list of social work journals was guided by the work of Hodge and Lacasse (2011). More specifically, Hodge and Lacasse (2011) identified 84 disciplinary social work journals based on a variety of sources, including *An Author's Guide to Social Work Journals* (NASW Press, 1997), Thyer's (2005) more recent listing of social work periodicals, and Genamics JournalSeek (<http://journalseek.net/>). Hodge and Lacasse (2011) examined the mission and aims of each journal, and eliminated journals that were specific to another field or had an

had an inter-disciplinary focus. For the current study, our primary search query included all journal titles in this list.

To help ensure comprehensive coverage of all possible core social work journals, we created a supplemental search query that extracted journal titles with any of the following terms:

- “social work”
- “social welfare”
- “social casework”
- “social service”
- “human service”
- “social development”
- “social environment”

Using this search queries, we extract journal article records from three major databases on the EbscoHost platform: PsycINFO, Social Service Abstracts, and Social Work Abstracts. Because of known gaps in indexing in databases (Holden, Barker, Kuppens, Rosenberg & LeBreton, 2014; Holden, Barker, Covert-Vail, Rosenberg, & Cohen, 2009), we also searched ProQuest, which linked to an additional 46 minor databases, including (but not limited to): ERIC, Sociological Abstracts and Worldwide Political Science Abstracts. In every search, we used a filter in attempt to extract only article records that were classified as *journal articles*, thereby excluding other forms of scholarly communication that was not relevant to the current study (e.g., book reviews, editorials, obituaries, etc.). We limited the timeframe to a 25 year period, from 1989 to 2013 (inclusive). Articles published in years 2014 and 2015 were excluded because of delays in indexing. Search results were exported in batches of article records based on the restrictions of the platform. Article records were text files in a *generic bibliographic format*. The article records contained various meta-data based on the database from which the article record was extracted. Meta-data include (but are not limited to): article title, journal title, publication year, author name(s), author affiliation(s), abstract, keywords, methodological classification, funding source, location of study, subject groups, digital object identifier (DOI), number of references, number of pages, etc. These files were post-processed into a structured database using a set of scripts written in the R statistical programming language. The initial search resulted in `r pretty(nrow(article.count.initial))` article records from `r journal.count.initial`.

The search queries were purposefully specified to be overly inclusive to ensure full coverage of all possible social work journals contained in the major and minor databases. Additionally, a visual inspection of the article records revealed problems in the original indexing of articles that requires subsequent data cleaning. Thus, our data cleaning procedures focused separately on journal titles and article records.

### **Data cleaning: Journal titles**

The first step of data cleaning involved fixing journal title names due to discrepant errors in indexing. For example, *Journal of Gay and Lesbian Social Services* was indexed as a journal separate from *Journal of Gay & Lesbian Social Services* (use of “and” vs. “&”). Other examples involved journals indexed with and without subtitles, or journals with and without the word “The” at the beginning of the title. Additionally, some journals changed titles over their history. For example, *Journal of Technology in Human Services* is formerly known as *Computers in Human Services*. In these situations, the former titles were merged with the current titles.

Many journals from allied health disciplines used one of the supplemental search terms in special editions, which were also included in the journal title. Thus, many non-social work journals were captured in the extraction process. To resolve this issue, we created a list of all journal titles that were not part of the core list defined by Hodge and Lacasse (2011). Study authors reviewed these journal titles and discussed whether each candidate title should be retained or excluded. When disagreements occurred, the study authors reviewed the mission and aims of the journals, names of editorial board members, and focus of the articles. A consensus was reached on all journal titles to be excluded and retained. After these procedures, the number of

article records and journal titles was reduced to `r pretty(nrow(article.count.journalsCleaned))` and `r journal.count.journalsCleaned`.

## Data cleaning: Article records

Our search procedures involved the use of three major databases and 46 minor databases, which resulted in some duplication of article records. Thus, a matching algorithm was constructed to identify and remove all duplicate article records. As previously noted, the filtering of only journal articles was not successful due to errors in the original indexing. Thus, other scholarly communications were captured in the extraction process. These other scholarly communications were readily identified due to specific patterns in the titles, such as the terms “Book Review,” “From the Editor,” and “Obituary.” Given the number of article records in the database, it was not feasible to manually extract these records. Instead, we created a separate database with representative examples of article records that we wanted to extract. We also added representative examples of article records to be retained. We then wrote a series of *regular expressions* to extract the problematic article records. Regular expressions are a sequence of characters that form search patterns, allowing extraction of records that match the specified pattern. In other words, it is a more sophisticated implementation of the search function common to all major word processors. We tested our regular expressions on the test database, achieving > 95% accuracy with regard to retaining and extracting article records. These regular expressions were then applied to the full database. We then excluded article records if they were missing essential meta-data, including author name(s), journal title, article title, and publication field. As an additional requirement, we required all articles to be  $\geq 3$  pages in length. Finally, we eliminated all social work journals with total article counts of < 10 for the entire 25 year window (*Issues in Social Work Education*, *Maatskaplike Werk/Social Work*, *Pediatric Social Work*, and *Critical Social Work*). These cleaning procedures resulted in a final database of 32,897 articles and 79 journals.

## Data quality checks

We performed a variety of checks on the final database. Because the size of the database, it was not feasible to manually inspect all records. Thus, we randomly sampled article records from the database and cross-checked the data we procured with the information listed with the journal’s homepage. We also inspected a scatter plot of the number of article records for each journal by the number of years the journal appeared within the 25 window timeframe. Journals that exhibited significant deviation from the regression line were inspected to ensure that our cleaning procedures did not systematically exclude article records or retain other scholarly communications. All observed discrepancies were verified as differences in journals actual publication output. Finally, we also made checks of number of journal article records for various journals indexed in the *Web of Science*. Although the counts were close – i.e., within 10% – similar indexing problems observed from our record extraction were also observed in the *Web of Science* database, such as misclassification of book reviews as journal articles. Thus, we are unable to quantify the actual amount of error in our database is due to problems in indexing. At the same time, we are relying on the same source of data that social work researchers use to inform their work. Too much cleaning and post-processing of the search results can give rise to validity issues because social work researchers are using an information source that contain the errors we have attempted to eliminate. The issue of reliability and validity are given further attention in the discussion section of this study.

## Analyses

# Results

## Aggregated summary (Figure 1a. 1b. – line charts)

Figures 1a and 1b are line charts showing the number of unique journal titles and journal articles published each year during the study time frame. In 1989, XX unique journals published a XX articles. The number

of journal titles showed some variability from 1989 to 1996, followed by nine years of steady growth. By 2005, the number of journal titles and journal articles being published was more than double than what was observed in 1989. The years 2005 to 2013 exhibited some variability in the number of titles and articles published. The maximum number of journal titles was observed in 2011 ( $n = XX$ ) and dropped to  $XX$  by 2013. The maximum number of journal articles published was in 2012, and this dropped to  $XX$  in 2013. The cumulative number of journal articles published over the study time frame was  $XX,XXX$ .

## Dis-aggregated summary (Figure 2. Small multiples)

Figure 2 is a small multiples plot, also referred to as a trellis or lattice chart, that dis-aggregates Figures 1a and 1b. More specifically, the small multiples plot shows the number of articles published each year for each journal throughout the study time frame (orange line). We also included data from Scopus for comparative purposes (purple line). The small multiples are sorted in descending order based on the total number of article records in the HD. The presentation of the sort is by row, so the overall plot is read from left to right. Abbreviations are used in Figure 2, and the full journal name is presented in Appendix A.

The small multiples plot is particularly useful to show the variability in publishing for each journal. In fact, according to Tufte (xxxx), small multiples are the best design solution for visually enforcing comparisons of change. The plot shows that *Families in Society*, *Social Work*, and *British Journal of Social Work* published the greatest number of articles throughout the study time frame. However, the number of articles published by *Families in Society* and *Social Work* appear to be decreasing, while other journals show an increase – e.g., *Research on Social Work Practice*, *International Social Work*, and *Journal of Human Behavior and the Social Environment*.

Data from Scopus was used to help show the overall coverage of the HD, in addition to revealing possible gaps in indexing. The HD and Scopus plotted values do not show exact alignment for a couple reasons. More specifically, the HD contains article records going back to 1989, whereas Scopus provides data starting from 1996. Additionally, the overall journal coverage of HD is much greater than Scopus – that is, Scopus provides data for only  $xx\%$  of the journal titles contained in the HD.

The Scopus data that are available show a high degree of correspondence with the HD. Small discrepancies between the Scopus and HD can be attributed to differences in indexing. That is, the article record count of Scopus includes editorials and book reviews, whereas these were excluded from the HD. In the construction of the HD, we applied a series of cleaning procedures (refer to Methods section) that further reduced the article count. Because Scopus data were available only in aggregate form, it was not possible to extract these data to make more refined comparisons.

And, finally, the small multiples plot is especially useful for identifying potential gaps in indexing, which is a problem that has been previously identified with respect to social work journals (see Holden; Holden). For example, Scopus appears to be missing article records for the *Journal of Gerontological Social Work* from  $XX$  to  $XX$ . And, the HD is missing records for various journals including (but not limited to) the *Journal of Ethnic and Cultural Diversity in Social Work*, *Journal of Social Work Disability & Rehabilitation*.

## Article attributes

Figures 3a and 3b provide summaries of two important article attributes – that is, the number of authors per article and the length of each article. Throughout the history of social work research, at least half of all social work articles were sole authored until 2008, as indicated by the median value in Figure 3a. Since 2008, at least half of all articles had two authors. Throughout the study time frame, the number of authors has increased steadily, as shown by the increasing mean value and its standard deviation. In 2013, the mean number of authors was approximately 2.25, with a standard deviation of 1.4. The length of the articles increased steadily from approximately 12 pages in 1989 to 15 pages in 2004. The last three years of the study time frame shows a fairly stable publishing trend, with half of all articles being 15 pages (Mean =  $XX$ , sd =  $XX$ ).



## Abstract analysis

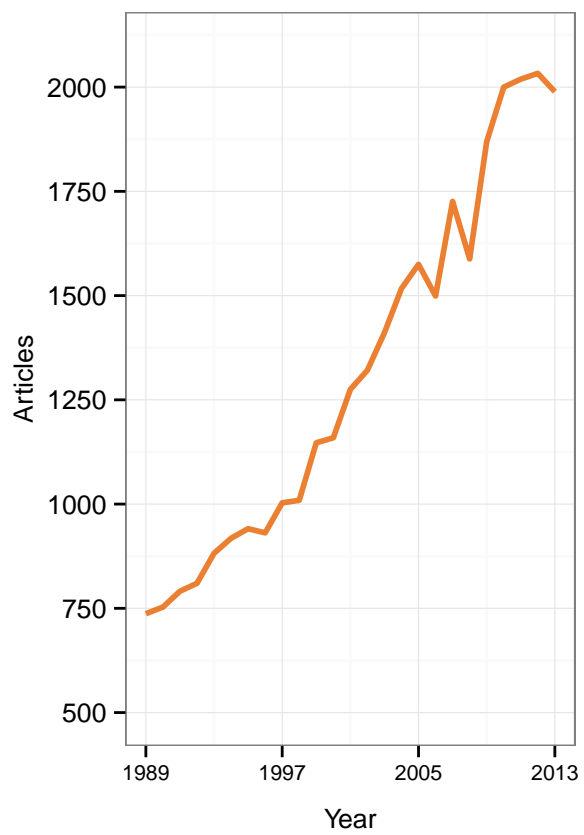
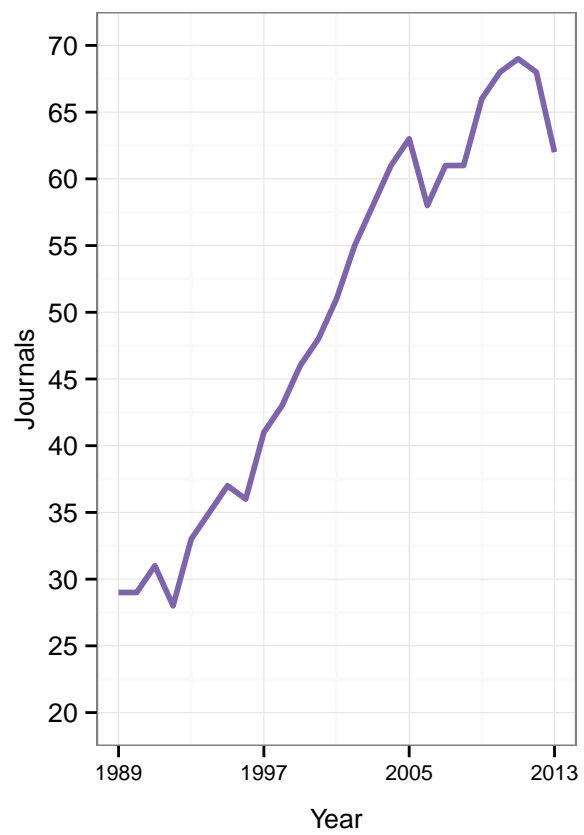
The final set of analyses involved text analysis of the article abstracts to determine whether trends in publishing could be readily identified based on the frequencies of specific words contained in article abstracts. In this analysis, we selected words that represented different *levels of evidence* within the evidence-based practices framework. EBP was selected as the focus of analysis, given that the topic area is a relatively new in social work, compared to other topics like child welfare and case management. Moreover, because EBP has received considerable attention in various areas of social work research, it was expected that the different *levels of evidence* would show both absolute and relative changes in their occurrences.

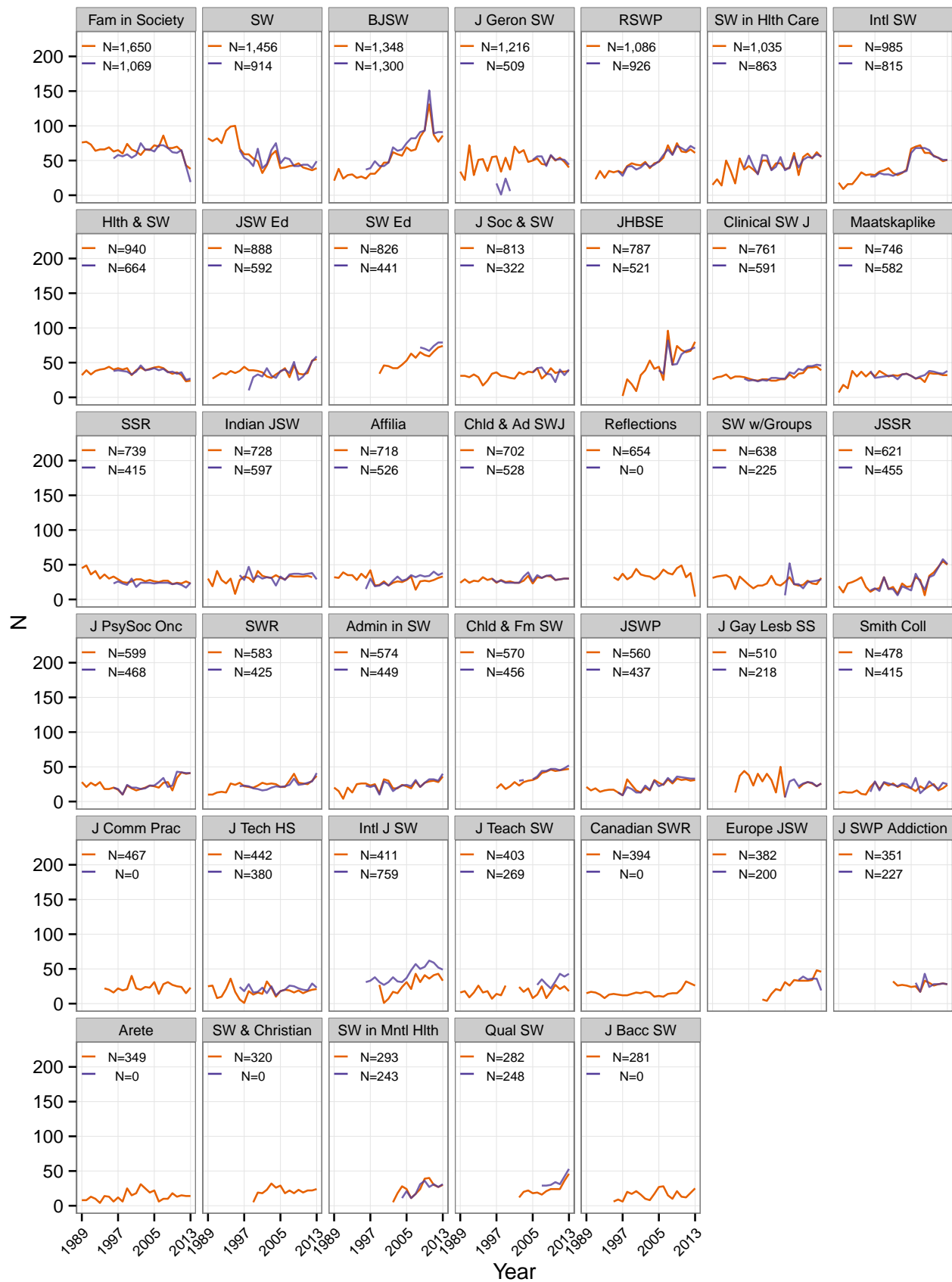
In this analysis, we focused on four different *levels of evidence*: Meta-analysis, systematic reviews, randomized controlled trials (RCTs), and quasi-experimental designs. To identify these key words in the article abstracts, we created a set of regular expressions of common spellings and synonyms for each type of evidence (see Appendix B). Regular expressions, also referred to as regex and regexp, are sequences of characters that define specific patterns in text documents and work much like the *find* functions common to word processors. The regular expressions used to define the *levels of evidence* are presented in Appendix B and the study's source code.

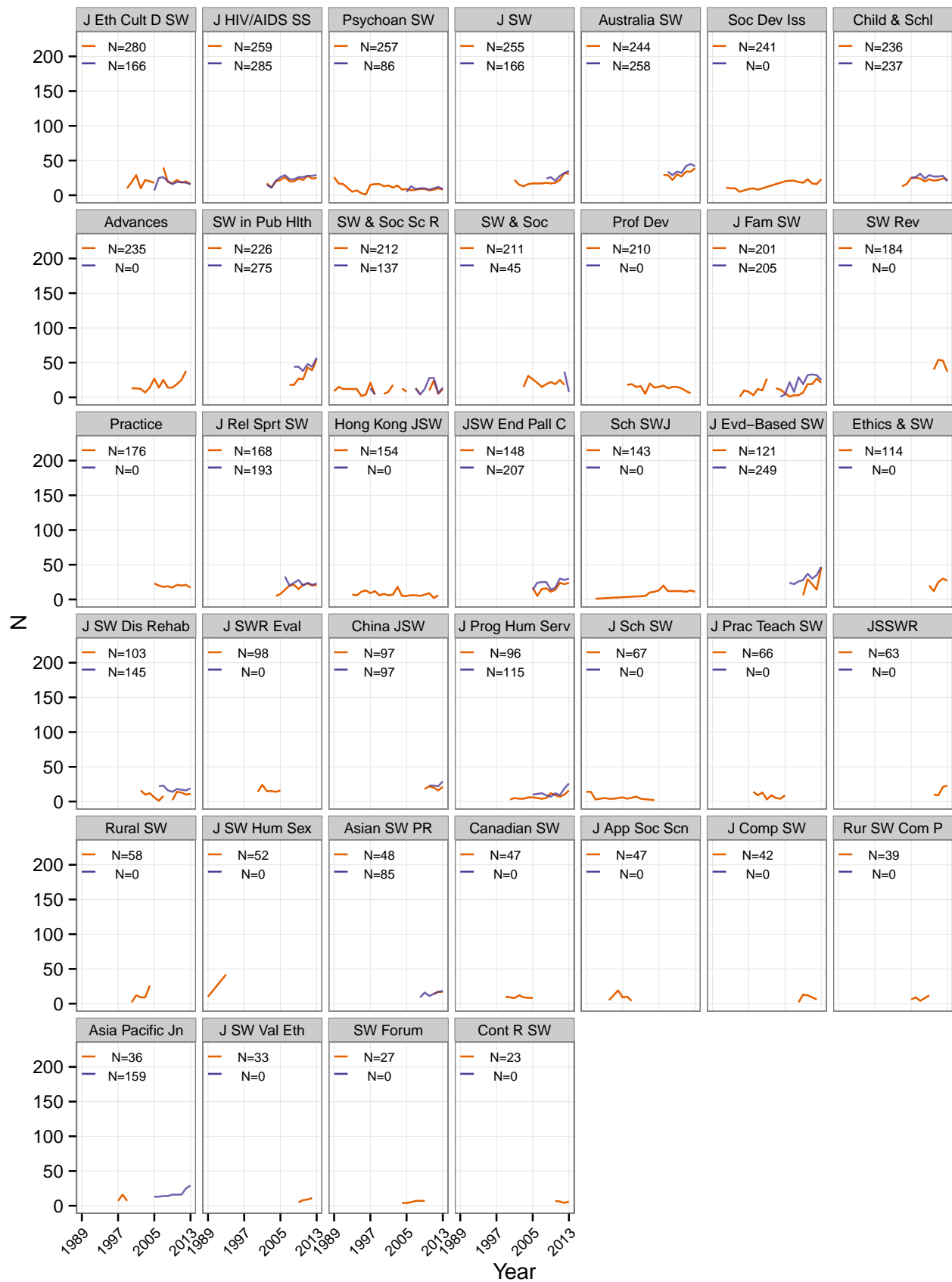
A stacked line chart was used to show absolute change in the number of occurrences for each level of evidence in article abstracts. The line chart shows increased numbers of occurrences for each level of evidence, with the greatest amount of growth starting in year 2008. The presence of systematic reviews and meta-analyses did not have a sustained presence in the literature until approximately 1995. In fact, the first occurrence of the key word *systematic review* did not appear until 1997 (cite).

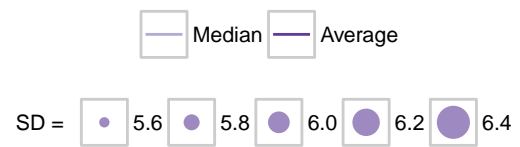
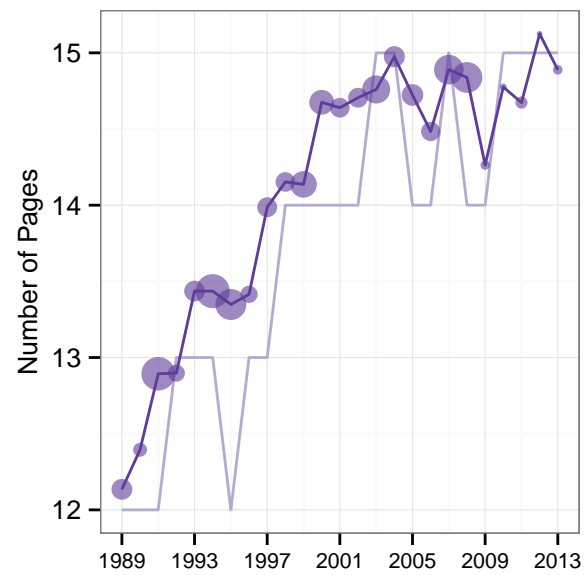
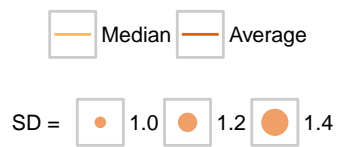
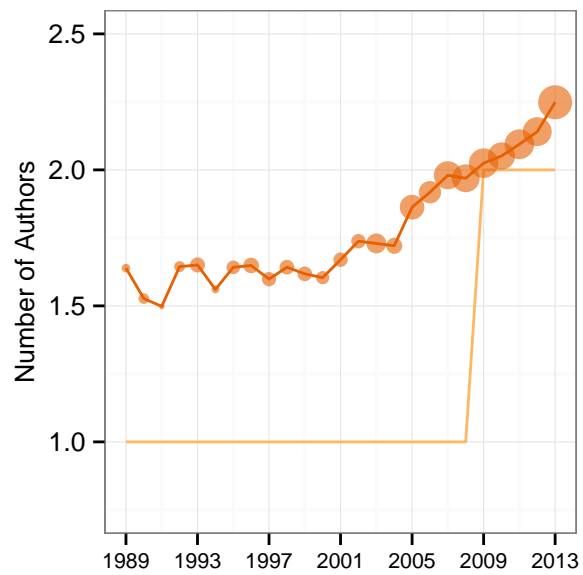
To show relative change, we plotted the proportion of overall articles published that contained the target key words for each year. In 1989, RCTs and Quasi-experimental studies were observed in approximately 1/2% of published articles. By 2013, the percentage of RCTs tripled and the percentage of quasi-experimental doubled. Meta-analyses appeared in 1989 and grew to almost 1/2% by 2013. Systematic reviews first appeared in 1997, then again in 2003, and grew to approximately 1% in 2013.

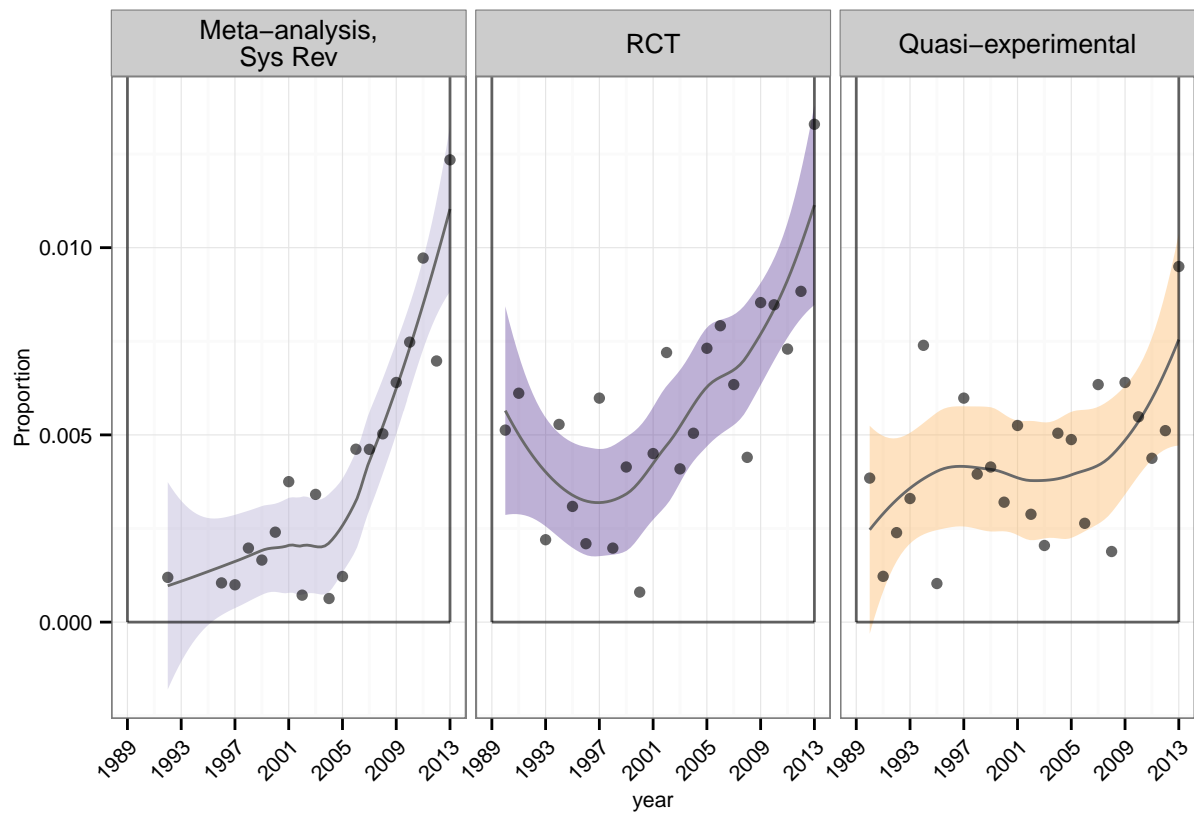
## Pareto chart

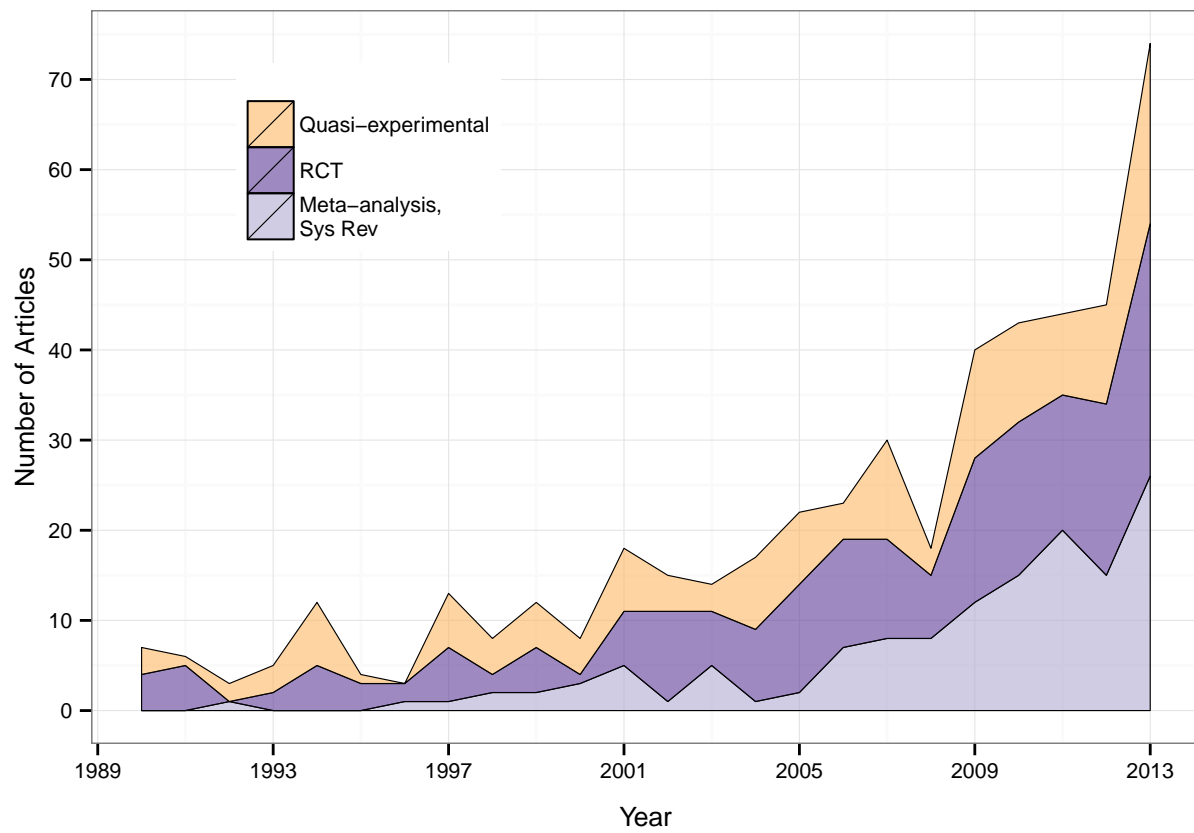


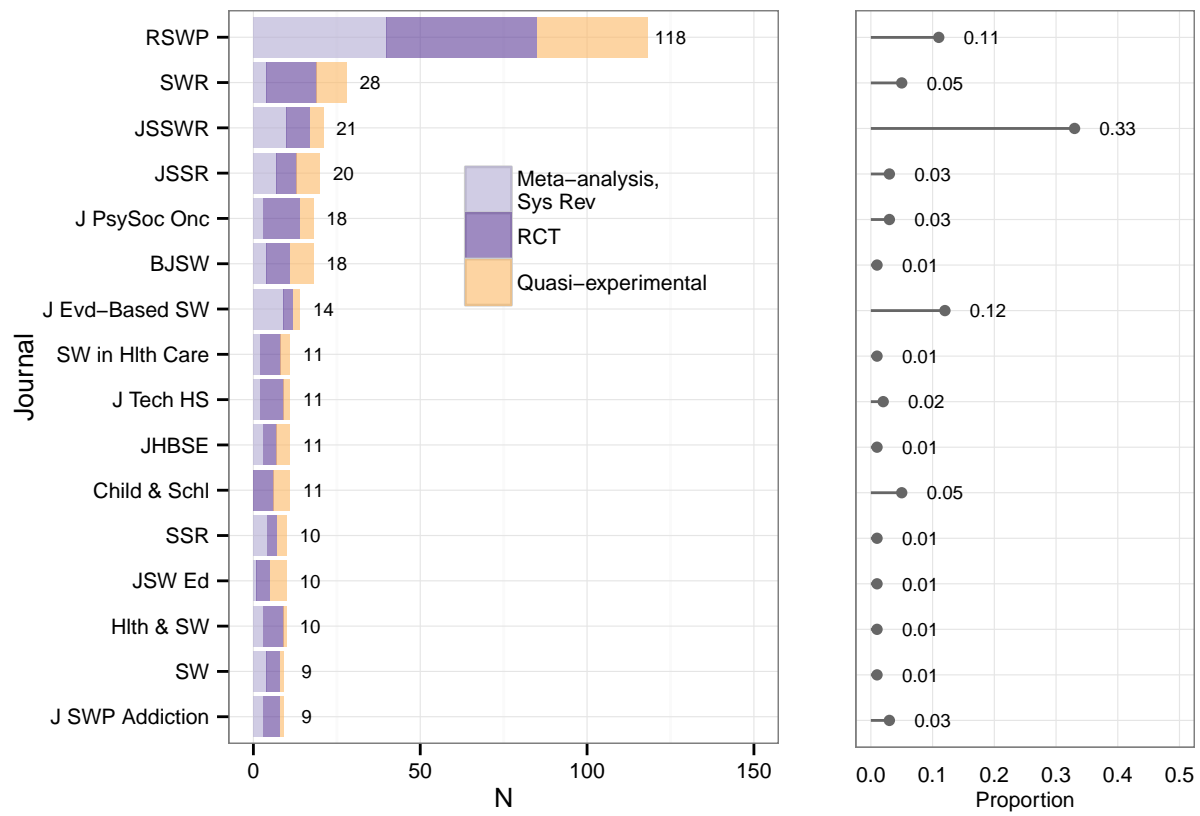














## Appendix A

Articles on Hodge and Lacasse (2011) but not in the current study.

- [1] "Annual of Social Work"
- [2] "Black Caucus"
- [3] "Caribbean Journal of Social Work"
- [4] "Critical Social Work"
- [5] "Electronic Journal of Social Work"
- [6] "IUC Journal of Social Work Theory & Practice"
- [7] "Japanese Journal of Social Services"
- [8] "Journal of Changsha Social Work"
- [9] "Journal of Forensic Social Work"
- [10] "Journal of Rural Social Work & Social Development"
- [11] "Journal of Social Work in Long-Term Care"
- [12] "The New Social Worker"
- [13] "The Social Worker/Le Travaillleur Social"
- [14] "The Spirituality & Social Work Forum"

Source: local data frame [4 x 2]

	record N
1	Critical Social Work 1
2	Issues in Social Work Education 1
3	Social Work/Maatskaplike Werk 1
4	Iowa Journal of School Social Work 2