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

May 26, 2015

Methods

Identification of social work journals

The primary unit of analysis for this study was the bibliographic records of social work journal articles (hereafter referred to as `article records`). The first step in the extraction process involved constructing a comprehensive list of social work journal titles. We used the work of Hodge and Lacasse (2011) as our base list. Hodge and Lacasse (2011) compiled a list of 84 unique social work journal titles using a variety of sources, includingt the book *An Author's Guide to Social Work Journals* (NASW Press, 1997), Thyer's (2005) more recent listing of social work periodicals, and the website *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 inter-disciplinary focus. All social work journals identified by Hodge and Lacasse (2011) were included on our base list.

To ensure comprehensive coverage, we supplemented our search by searching the Internet and bibliogpraphic databases that contained any of the following terms:

- "social work"
- "social welfare"
- "social casework"
- "social service"
- "human service"
- "social development"
- "social environment"

Journal titles that were identified with any of these terms were reviewed by the study team to determine whether they should be included in the study. When disagreements occurred, the authors reviewed

the mission and aims of the journals, names of editorial board members, and content of representative articles. A consensus was reached on all journal titles to be excluded and retained. In this process, an additional XX journal were identified, producing a final target list of XX journals.

Extraction of article records

After finalizing the target list of social work journals, we extracted article records from three major databases on the EbscoHost platform (PsycINFO, Social Science Abstracts, and Social Work Abstracts) and xx minor databases on ProQuest (e.g., ERIC, Sociological Abstracts, Worldwide Political Science Abstracts). Using the `publication source` fields of the search interface, we searched for article records that contained the names of each social work journal. Searches were limited to a quarter century time-frame, 1989 to 2013. We excluded years 2014 and 2015 due to delays in indexing of records that were observed. In the extraction process, we also limited the records to only those classified as "journal articles," thereby excluding other types of scholarly communications, such as editorials, comments on articles, letters to the editor, book reviews, and obituaries.

Search results were saved and exported as plain text files in *generic bibliographic format*. The article records contained meta-data, including (but 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. It should be noted that the databases provided different pieces of meta-data, or meta-data in varying formats. Thus, extensive post-process of the article records was necessary to make the article records comparable and analyzable.

Post-processing of article records

Articles were post-processed from the *generic bibliographic format* to a structured database with computer scripts authored specifically for this study using the programming language R. Post-processing of the files involved addressing numerous problems that occurred during the original indexing of the article records. For example, journal titles that included the word `and'' sometimes used an amersand (&). Other examples included inconsistent use of the word`The" at the beginning of the title, or the inconsistent use of subtitles. It should be noted that some journals changed titles over their publication history. For example, *Journal of Technology in Human Services* is formerly known as *Computers in Human Services*.

All historical titles were merged with their current titles and analyzed as a single journal in this study. The source code provides complete details of all the procedures to resolve discrepancies and historical titles.

In addition to the foregoing date quality problem, a variety of other problems were identified and resolved. For example, the filtering of article records did not successfully exclude all other types of documents (e.g., editorials, book reviews, and obituaries). Thus, we used a set of text matching procedures called *regular expressions* to identify other documents that were incorrectly classified as journal articles. We observed $>$ 95% accuracy identifying non-article records using a separate database designed specifically for script testing. We excluded article records if they were missing essential meta-data for this study, including author name(s), journal title, article title, and year of publication. Finally, we eliminated all social work journals with total article counts of < 10 for the entire 25 year window (*Issues in Social Work Education*, *Pediatric Social Work*, and *Critical Social Work*).

Analytic strategy

We conducted a series of descriptive analyses using various types of analyses and meta-data to demonstrate the utility of the database. Our first set of analyses focused on describing the overall *amount* and *growth* of social work research by summarizing the number of unique journal titles and articles published each year throughout the 25-year time frame. We then disaggregated these data by examining the number of articles published by each journal. We also performed an additional check on the quality of the data by making systematic comparisons with aggregate publication data from Scopus. Scopus data was selected because it is the single largest abstract and citation database of scientific journals and was not linked to either the EbscoHost or ProQuest databases in the record extraction process.

After describing the *amount* and *growth* of social work research, we conducted a series of analyses to test the utility of using meta=data to reveal publication trends. As an indicator of the growth of *team science* we summarize counts of the number of authors per article over time. And, in light of the shift from print to electronic articles, we examined whether the length of articles have changed during this 25-year time frame. Finally, we demonstrate the use of full-text abstracts as the basis for topical trend analysis. For purposes of demonstration, we use the concept of "evidence-based practices" (and related concepts) to show changes in topics and areas of foci in the social work research.

Reproducible research

In addition to post-processing original article records, R was also used for all other data analysis. R is an ideal analytic tool because it is open-source and has a suite of related tools for producing reproducible research. More specifically, all the source code is written using Rmarkdown – an open-source system that integrates data management, data analysis, and the written report. The source code and data extracts are also curated at openICPSR of the Inter-university Consotrium for Political and Social Research of the University of Michigan. The source code is also available from the first author's GitHub repository, where the code will be maintained and updated to support future research projects. It should be noted that the full data files cannot be made publicly available because these are considered proprietary data of the database services. However, to ensure full compliance with the best practices of reproducible research (cite King), we have provided the journal editor with the following conditions under which we would release the data: (1) the requester must be affilitated with an insitution that provides access to both the EbscoHost and ProQuest platforms along with the databases used in this study or (2) hold a personal subscription to each platform and the databases utilized here. This is designed to meet the standards of maximum allowable data transparency while respecting the proprietary nature of the data.

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 should both absolute and relative changes in their occurrences.

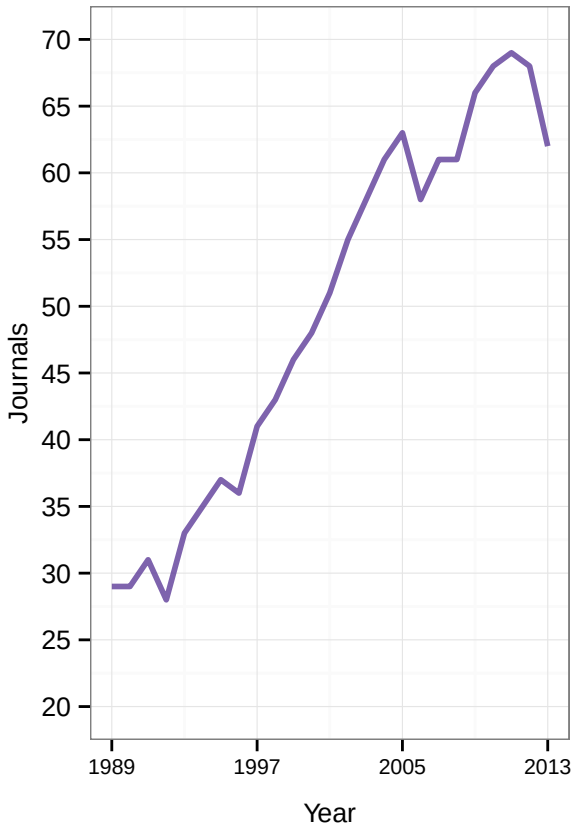
In this analysis, we focused on four different *levels of evidence*: Meta-analysis, systematic reviews, randomized controlled trials (RCTs), and quasi-experimental designs. To identify these key words in the article abstracts, we created a set of regular expressions of common spellings and synonyms for 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.

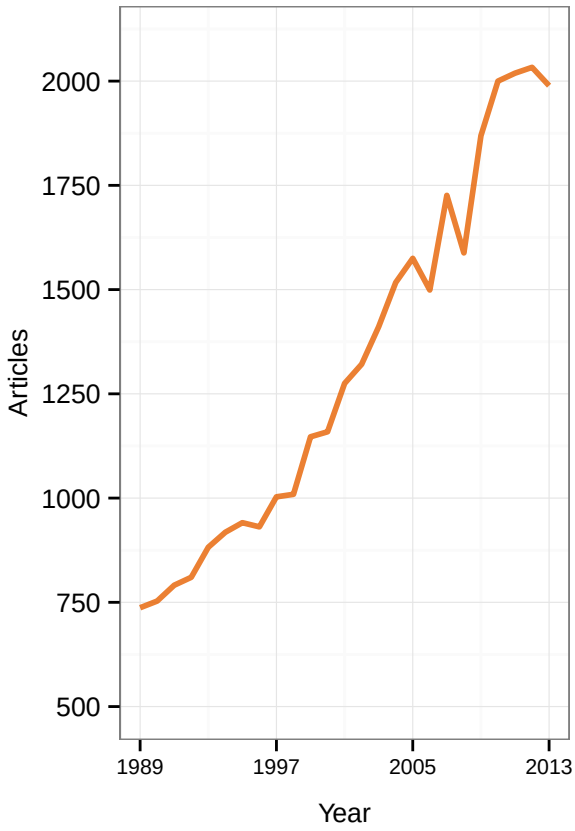
A stacked line chart was used to show absolute change in the number of occurrences for each level of evidence in article abstracts. The line chart shows increased numbers of occurrences for each level of evidence, with the greatest amount of growth starting in year 2008. The presence of systematic reviews 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).

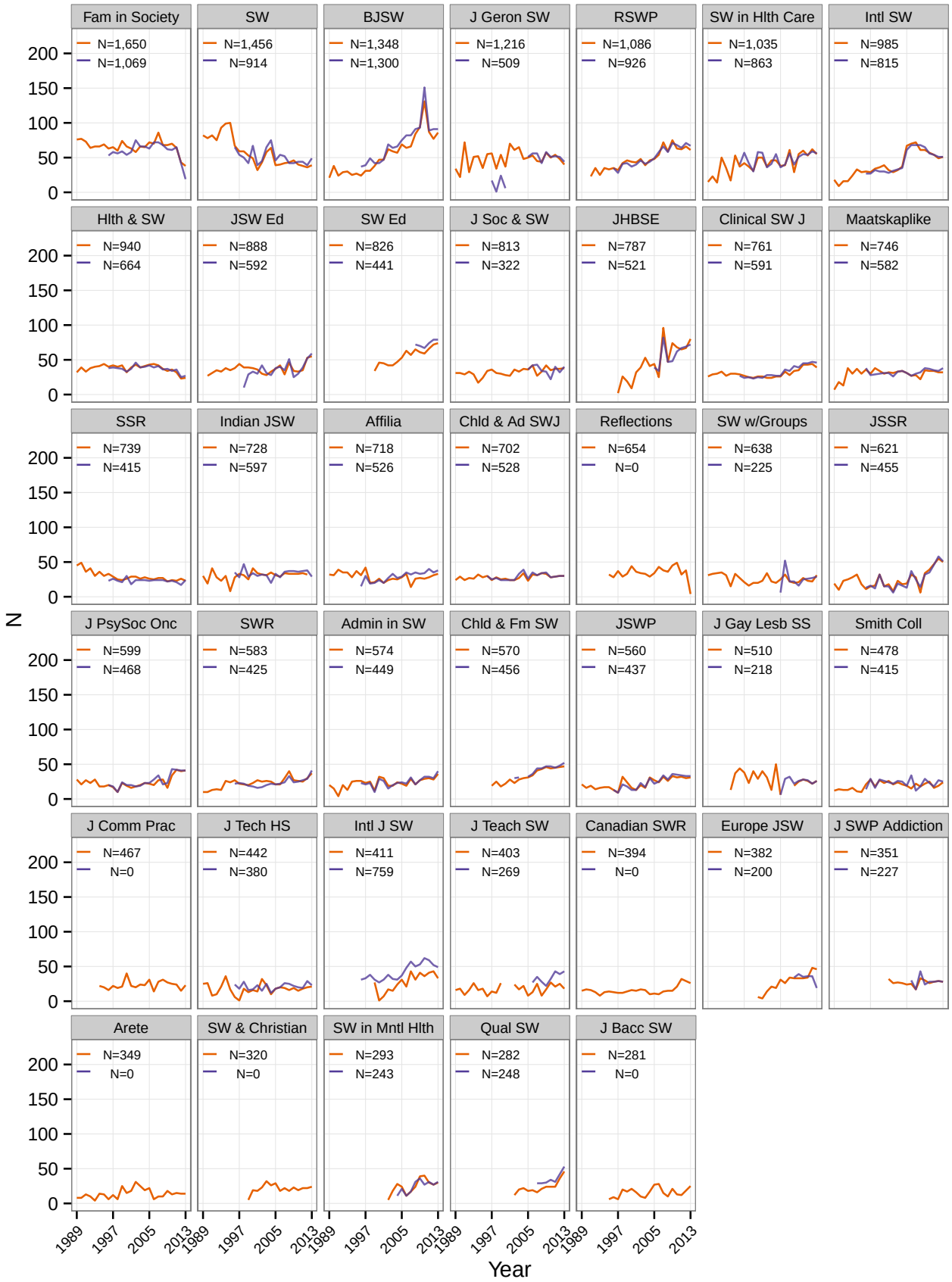
To show relative change, we plotted the proportion of overall articles published that contained the target key words for each year. In 1989, RCTs and Quasi-experimental studies were observed in approximately 1/2% of published articles. By 2013, the percentage of RCTs tripled and the percentage of quasi-experimental

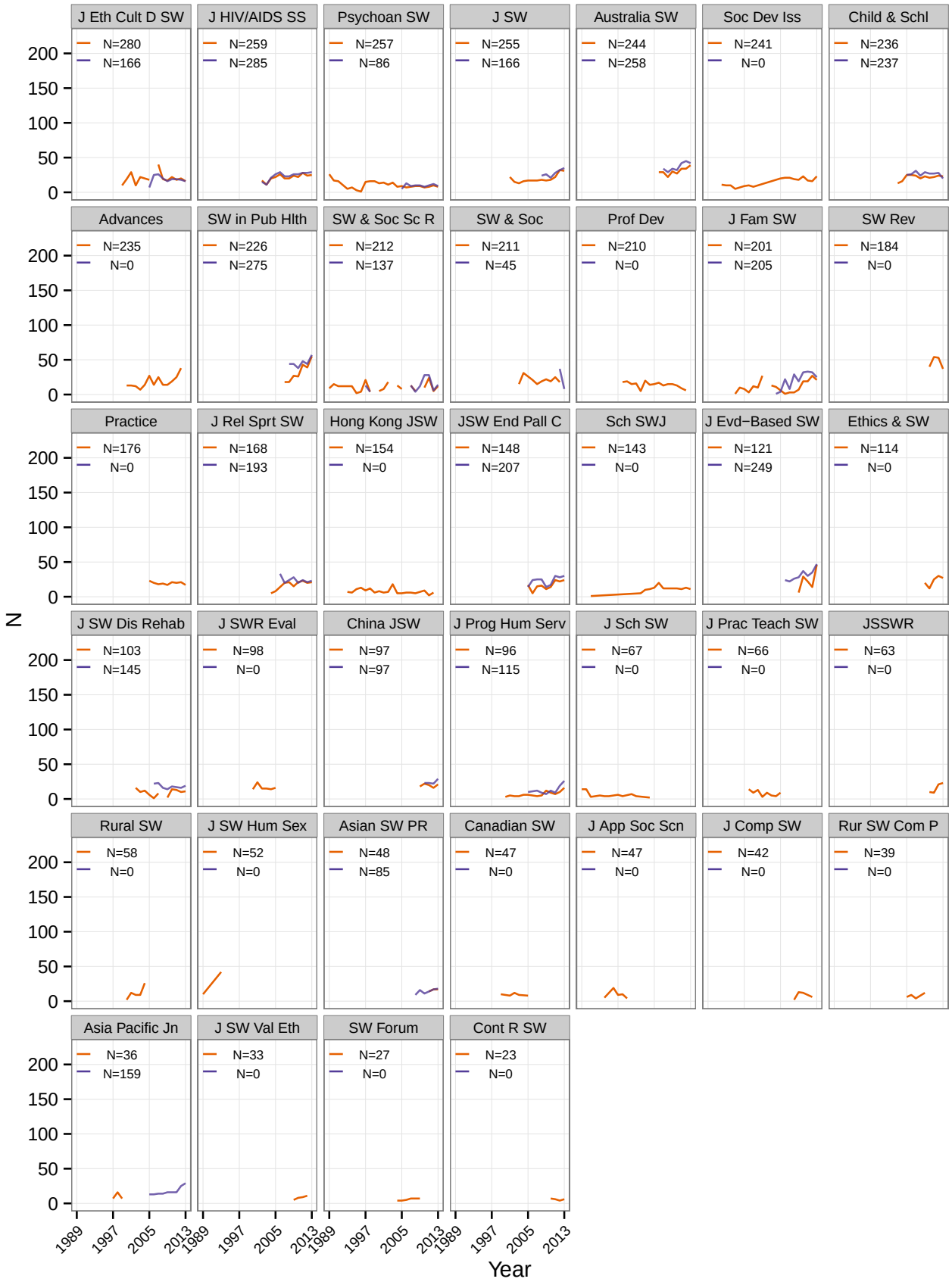
doubled. Meta-analyses appeared in 1989 and grew to almost 1/2% by 2013. Systematic reviews first appeared in 1997, then again in 2003, and grew to approximately 1% in 2013.

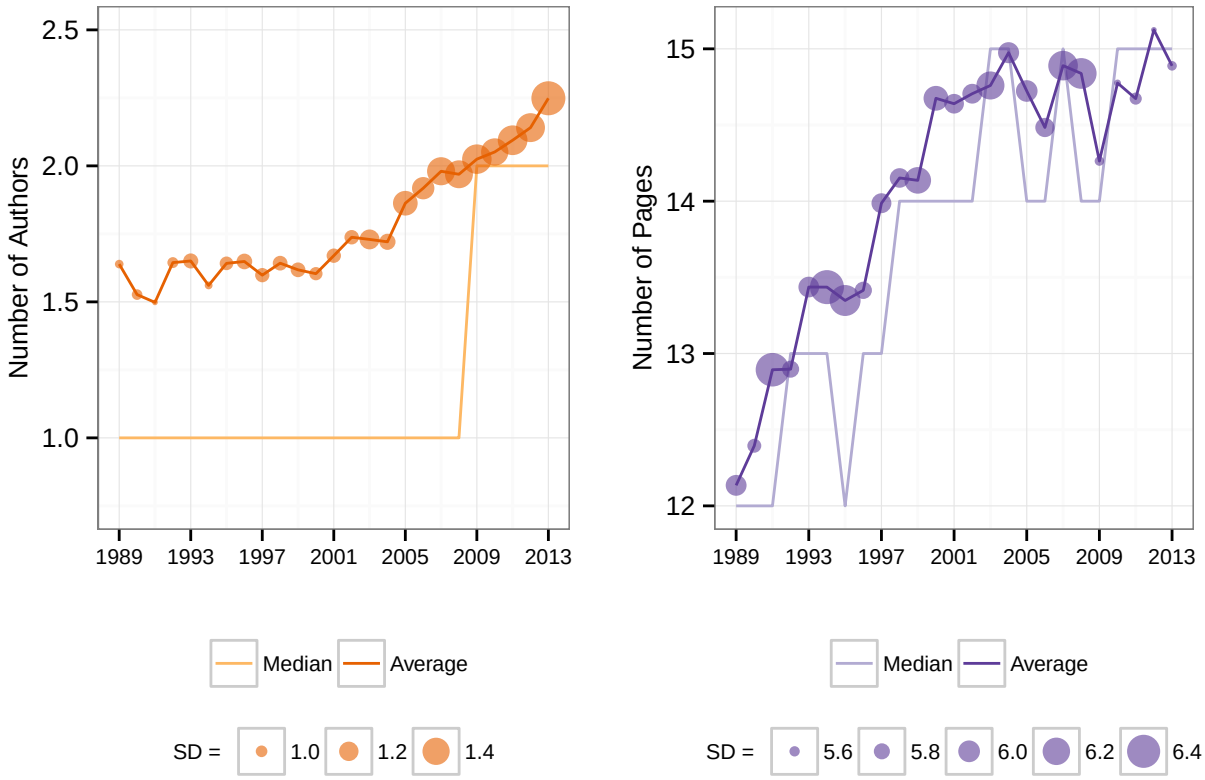
Pareto chart

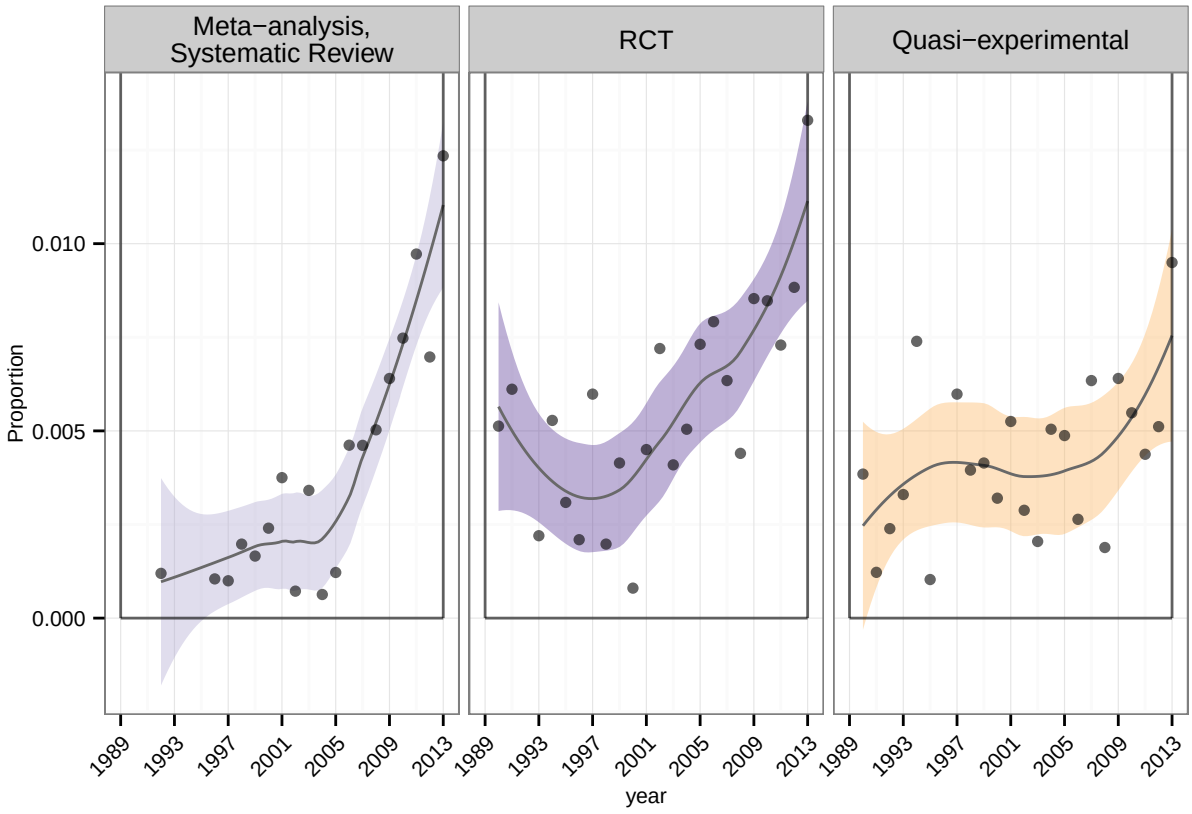

Journals
70
65
60
55
50
45
40
35
30
25
20
1989
1997
2005
2013
Year

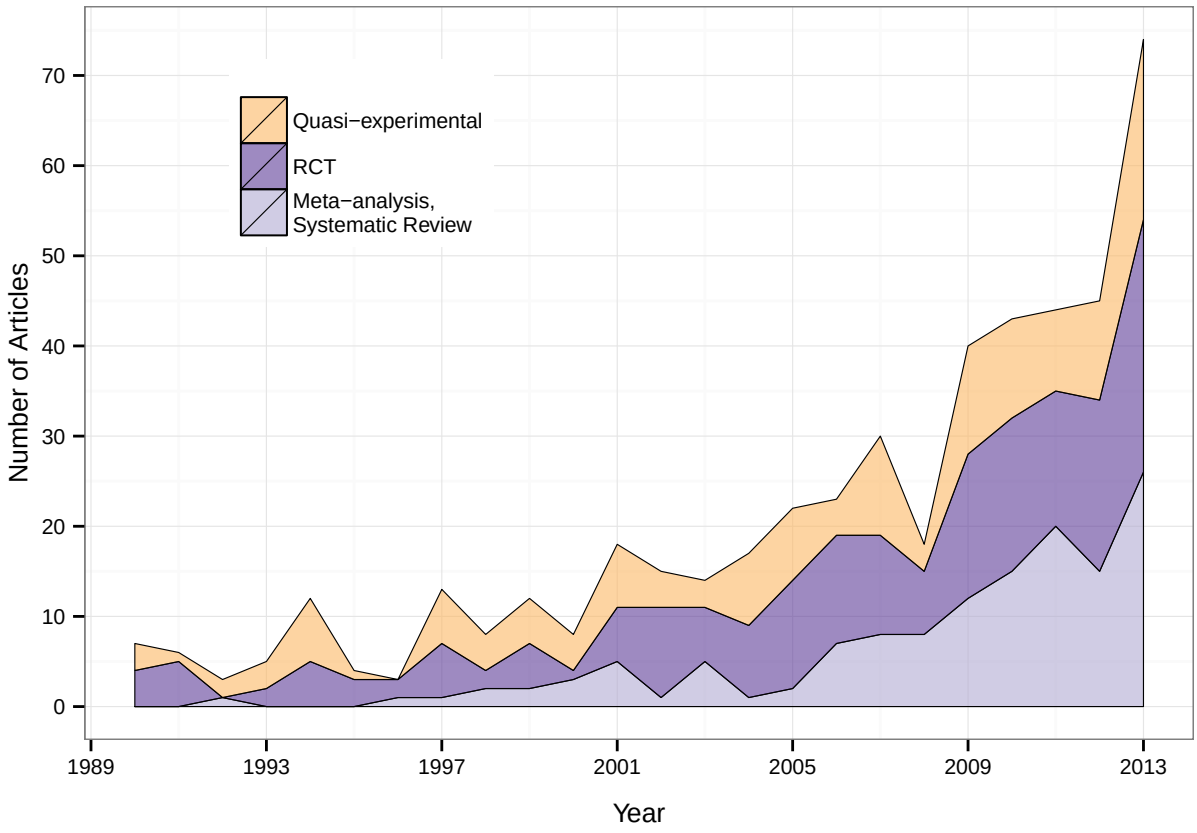

Articles
2000
1750
1500
1250
1000
750
500
1989
1997
2005
2013
Year

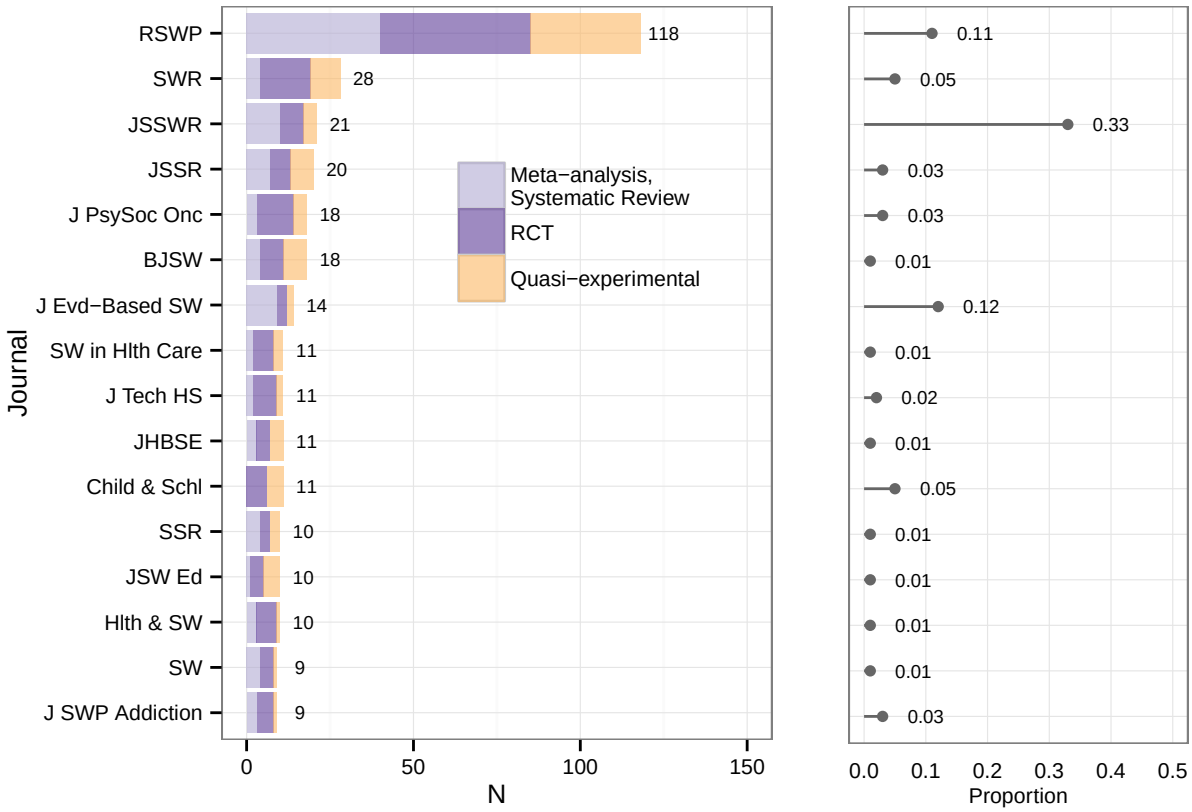

Fam in Society
N=1,650
N=1,069
SW
N=1,456
N=914
BJSW
N=1,348
N=1,300
J Geron SW
N=1,216
N=509
RSWP
N=1,086
N=926
SW in Hlth Care
N=1,035
N=863
Intl SW
N=985
N=815
Hlth & SW
N=940
N=664
JSW Ed
N=888
N=592
SW Ed
N=826
N=441
J Soc & SW
N=813
N=322
JHBSE
N=787
N=521
Clinical SW J
N=761
N=591
Maatskaplike
N=746
N=582
SSR
N=739
N=415
Indian JSW
N=728
N=597
Affilia
N=718
N=526
Chld & Ad SWJ
N=702
N=528
Reflections
N=654
N=0
SW w/Groups
N=638
N=225
JSSR
N=621
N=455
J PsySoc Onc
N=599
N=468
SWR
N=583
N=425
Admin in SW
N=574
N=449
Chld & Fm SW
N=570
N=456
JSWP
N=560
N=437
J Gay Lesb SS
N=510
N=218
Smith Coll
N=478
N=415
J Comm Prac
N=467
N=0
J Tech HS
N=442
N=380
Intl J SW
N=411
N=759
J Teach SW
N=403
N=269
Canadian SWR
N=394
N=0
Europe JSW
N=382
N=200
J SWP Addiction
N=351
N=227
Arete
N=349
N=0
SW & Christian
N=320
N=0
SW in Mntl Hlth
N=293
N=243
Qual SW
N=282
N=248
J Bacc SW
N=281
N=0
N
0
50
100
150
200
1989
1997
2005
2013
Year


J Eth Cult D SW
N=280
N=166
J HIV/AIDS SS
N=259
N=285
Psychoan SW
N=257
N=86
J SW
N=255
N=166
Australia SW
N=244
N=258
Soc Dev Iss
N=241
N=0
Child & Schl
N=236
N=237
Advances
N=235
N=0
SW in Pub Hlth
N=226
N=275
SW & Soc Sc R
N=212
N=137
SW & Soc
N=211
N=45
Prof Dev
N=210
N=0
J Fam SW
N=201
N=205
SW Rev
N=184
N=0
Practice
N=176
N=0
J Rel Sprt SW
N=168
N=193
Hong Kong JSW
N=154
N=0
JSW End Pall C
N=148
N=207
Sch SWJ
N=143
N=0
J Evd−Based SW
N=121
N=249
Ethics & SW
N=114
N=0
J SW Dis Rehab
N=103
N=145
J SWR Eval
N=98
N=0
China JSW
N=97
N=97
J Prog Hum Serv
N=96
N=115
J Sch SW
N=67
N=0
J Prac Teach SW
N=66
N=0
JSSWR
N=63
N=0
Rural SW
N=58
N=0
J SW Hum Sex
N=52
N=0
Asian SW PR
N=48
N=85
Canadian SW
N=47
N=0
J App Soc Scn
N=47
N=0
J Comp SW
N=42
N=0
Rur SW Com P
N=39
N=0
Asia Pacific Jn
N=36
N=159
J SW Val Eth
N=33
N=0
SW Forum
N=27
N=0
Cont R SW
N=23
N=0
N
0
50
100
150
200
1989
1997
2005
2013
Year


Number of Authors
2.5
2.0
1.5
1.0
1989
1993
1997
2001
2005
2009
2013
Median
Average
SD =
1.0
1.2
1.4
Number of Pages
15
14
13
12
1989
1993
1997
2001
2005
2009
2013
Median
Average
SD =
5.6
5.8
6.0
6.2
6.4


Meta−analysis,
Systematic Review
RCT
Quasi−experimental
Proportion
0.010
0.005
0.000
1989
1993
1997
2001
2005
2009
2013
year


Quasi−experimental
RCT
Meta−analysis,
Systematic Review
Number of Articles
0
10
20
30
40
50
60
70
1989
1993
1997
2001
2005
2009
2013
Year


RSWP
SWR
JSSWR
JSSR
J PsySoc Onc
BJSW
J Evd−Based SW
SW in Hlth Care
J Tech HS
JHBSE
Child & Schl
SSR
JSW Ed
Hlth & SW
SW
J SWP Addiction
Journal
118
28
21
20
18
18
14
11
11
11
11
10
10
10
9
9
Meta−analysis, Systematic Review
RCT
Quasi−experimental
0
50
100
150
N
0.11
0.05
0.33
0.03
0.03
0.01
0.12
0.01
0.02
0.01
0.05
0.01
0.01
0.01
0.01
0.03
0.0
0.1
0.2
0.3
0.4
0.5
Proportion

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 Travaileur 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
```