25 years of Social Work Research

Summary of Data Cleaning Procedures

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 recrods 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 1988 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 36,094 article records from 117 different journals.

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 Technnology 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 35,667 and 82 (respectively).

Data cleaning: Article records

Our search procedures involved the use of three major databased 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 extractio 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 feasbile to manually extract these recrods. 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 ≥ 3 pages in length. Finally, we eliminated all social work journals with total article counts of < 10 for the enitre 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 32,047 and 79.
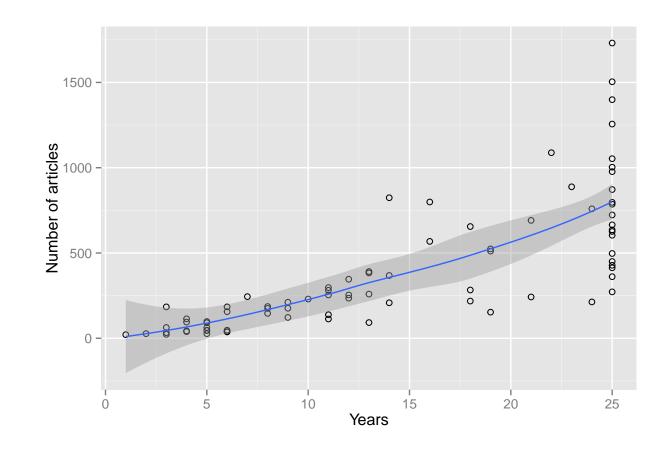
Data quality checks

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.

title	first	last	n	H
Administration in Social Work	1988	2013	604	Y
Advances in Social Work	2000	2012	235	Y
Affilia	2007	2013	184	Y
Arete	1988	2013	361	Y
Asia Pacific Journal of Social Work & Development	1997	2003	37	Y
Australian Social Work	2006	2013	244	Y
British Journal of Social Work	1988	2013	1399	Y
Canadian Social Work	1999	2005	47	Y
Canadian Social Work Review	1988	2013	413	Y
Child & Adolescent Social Work Journal	1988	2013	723	Y
Child & Family Social Work	1997	2013	568	Y
Children & Schools	2003	2013	230	Y
China Journal of Social Work	2009	2013	95	Y
Clinical Social Work Journal	1988	2013	786	Y
Contemporary Rural Social Work	2010	2013	23	N
Ethics & Social Welfare	2009	2013	114	N
European Journal of Social Work	2000	2013	391	Y
Families in Society	1988	2013	1731	Y
Health & Social Work	1988	2013	977	Y
Indian Journal of Social Work	1988	2012	759	Y
International Journal of Social Welfare	2000	2013	383	Y
International Social Work	1988	2013	1003	Y
Journal of Applied Social Sciences	1994	1999	47	Y
Journal of Baccalaureate Social Work	1995	2013	283	Y
Journal of Comparative Social Welfare	2008	2012	43	Y
Journal of Ethnic & Cultural Diversity in Social Work	2000	2013	259	Y
Journal of Evidence-Based Social Work	2004	2013	122	Y
Journal of Family Social Work	1995	2013	218	Y
Journal of Gay & Lesbian Social Services	1994	2013	511	Y
Journal of Gerontological Social Work	1988	2013	1256	Y
Journal of HIV/AIDS & Social Services	2002	2013	254	Y
Journal of Human Behavior in the Social Environment	1997	2013	799	Y
Journal of Multicultural Social Work	1999	2000	21	N
Journal of Practice Teaching in Social Work & Health	1999	2005	39	Y
Journal of Practice Teaching in Social Work & Practice	1998	2000	27	N
Journal of Progressive Human Services	2000	2013	92	Y
Journal of Psychosocial Oncology	1988	2013	626	Y
Journal of Religion & Spirituality in Social Work	2004	2013	176	Y
Journal of Social Service Research	1988	2013	635	Y
Journal of Social Work	2001	2013	254	Y
Journal of Social Work & Human Sexuality	1988	1993	45	N
Journal of Social Work Education	1990	2013	888	Y
Journal of Social Work Practice	1999	2013	368	Y
Journal of Social Work Practice in the Addictions	2001	2013	346	Y
Journal of Social Work Research & Evaluation	2000	2005	98	Y
Journal of Social Work Values & Ethics	2009	2012	33	Y
Journal of Social Work in Disability & Rehabilitation	2002	2013	113	Y
Journal of Social Work in End-of-Life & Palliative Care	2005	2013	146	Y
Journal of Sociology & Social Welfare	1988	2013	872	Y
Journal of Teaching in Social Work	1988	2013	428	Y
Journal of Technology in Human Services	1988	2013	450	Y
Journal of the Society for Social Work & Research	2010	2013	63	N
Practice: Social Work in Action	2005	2013	175	Y

```
                  Prevention in Human Services  1988 1993   91 N
                     Professional Development  1998 2012  208 Y
                  Psychoanalytic Social Work  1988 2013  272 Y
                    Qualitative Social Work  2002 2013  281 Y
    Reflections: Narratives of Professional Helping  1995 2013  655 Y
          Research on Social Work Practice  1991 2013 1088 Y
                        Rural Social Work  2000 2004   39 Y
               School Social Work Journal  2002 2013  138 Y
       Smith College Studies in Social Work  1988 2013  497 Y
                      Social Development  1992 2013  691 N
             Social Development Issues  1992 2013  242 Y
                 Social Service Review  1988 2013  797 Y
                          Social Work  1988 2013 1504 Y
             Social Work & Christianity  2005 2013  186 Y
    Social Work & Social Sciences Review  1989 2013  213 Y
                Social Work & Society  2003 2012  211 Y
                Social Work Education  1999 2013  824 Y
                   Social Work Forum  2004 2009   27 Y
               Social Work Research  1994 2013  524 Y
      Social Work Research & Abstracts  1988 1993   63 N
                  Social Work Review  2010 2013  184 Y
              Social Work With Groups  1988 2013  665 Y
           Social Work in Health Care  1988 2013 1053 Y
          Social Work in Mental Health  2002 2013  297 Y
          Social Work in Public Health  2007 2013  156 Y
    The Hong Kong Journal of Social Work  1993 2012  153 Y
```

Differences

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"
```