

## Hard or Soft Searching? Electronic Database Versus Hand Searching in Media Research

*Stephannie C. Roy, Guy Faulkner & Sara-Jane Finlay*

**Key words:**  
computerised  
indexes, hand-  
searching, media  
analysis, reliability

**Abstract:** It is important for qualitative media researchers to consider the impact of their research objectives on the sample frame imposed and subsequent data-collection methods. To illustrate this, we present some of the issues we encountered in determining a method of gathering physical activity articles in daily newspapers. We consider the implications of search choices for our sample, highlight the impact of using hardcopy hand-searches and electronic indexes and emphasise the importance of conducting a study to determine the reliability of hand-searching versus electronic index search methods. We suggest that researchers should be aware of the benefits and drawbacks of search methods including what kinds of information these methods yield and the possible effects on the research project. We conclude by highlighting the importance of these discussions to the reliability of content analysis.

### Table of Contents

- [1. Using Electronic Search Engines](#)
- [2. Researching Health and Physical Activity in the Canadian Mass Media](#)
  - [2.1 Locating relevant media texts](#)
    - [2.1.1 Sampling](#)
    - [2.1.2 Search and Retrieval](#)
  - [2.2 Search and retrieval comparison](#)
  - [2.3 Is more better? The implications of combining search strategies](#)
- [3. Conclusions](#)
- [Acknowledgements](#)
- [References](#)
- [Authors](#)
- [Citation](#)

## 1. Using Electronic Search Engines

There is a large and growing corpus of qualitative research examining media texts, however there has been little discussion within this literature of the searching and sampling procedures guiding this work. While most media content analyses provide some details of the sample size, inclusion and exclusion criteria, and inter-rater reliability, little discussion considers the development of the "meaning system" which is initially imposed on the method, that is, the determination of a category system used to decide which content will "fit" into an investigation (MCQUAIL, 2005). One of the most obvious places in which this initially occurs is in the determination of the sample and the processes used to search the sample. In this article, we take a step back from simply reporting an analysis to consider some of the issues we encountered in determining a method for selecting articles on health and physical activity which appeared in Canadian

daily newspapers. Like MANEY and OLIVER (2001) we suggest that it is necessary to include the details of selection structures to allow for comparison between studies and to recognise the potential implications for qualitative researchers of choosing different sampling and searching strategies. [1]

From November 2004 to April 2005, we analysed the content of a rolling sample of eleven daily newspapers and four daily national news broadcasts for stories which contained references to physical activity and health. While the television news programs were recorded and viewed; determining the most efficient means to search almost 50 newspapers a week was less clear. There are a number of strategies used to search media content such as hand searches of individual newspapers, the use of published indexes like *The Times Index*, *Canadian Periodical Index* or the *Reader's Guide to Periodical Literature*, or online searches through databases such as *Canadian Newsstand* or *Factiva* or individual CD/DVD newspaper indexes. [2]

In hand-searching for relevant newspaper articles, the researcher must read (or skim) every issue of the newspapers in the sample to collect articles that meet the search criteria. Hand-searching certainly permits the researcher to most closely approximate the actual reading processes of the average newspaper consumer, and is thorough; however it can also be time-consuming, messy, and tedious. [3]

In terms of published indexes, PEARSON and SOOTHILL (2003) found *The Times Index* very reliable for their examination of murders reported in *The Times*. However, as ZOLLARS (1994) explains, there are several methodological difficulties that may occur due to historical and idiosyncratic changes in the format of the index and the category headings, and she suggests researchers proceed cautiously when using print indexes. Likewise, time can also be a factor in the production of these indexes which are published several months after the newspapers appear. If information is needed quickly, this may cause delay. [4]

The final method is to conduct full-text keyword searches using internet or CD/DVD newspaper indexes and databases. It is likely that these types of searches are growing in favour because they allow the researcher to search the entire text of a newspaper for relevant keywords, meaning that every instance of the keyword can be found and ostensibly, all articles can be located. In addition, they allow the researcher the ease and comfort of conducting the search from their computer—one is not limited to locally available newspapers but can access all that are available in the database. Computerised newspaper searching, however, carries with it some methodological concerns, such as the identification of irrelevant items or false positives (SOOTHILL & GROVER, 1997). [5]

MANEY and OLIVER (2001, p. 132) looked at the implications of constructing events from different source materials, noting that "each record source and search strategy identifies a different subset of events that produce different views of 'what happened' in that month". While researchers have considered the use of various indexes to compile media samples (e.g. PEARSON & SOOTHILL, 2003;

SOOTHILL & GROVER, 1997; ZOLLARS, 1994), few provide an explanation of the methodological implications of the choices made by researchers in determining their sample in a timely and efficient manner or the impact of research-specific objectives on the sample frame. In this paper, we look at the impact of two key research objectives on the method used to identify our sample and the implications for the content of the sample itself. These objectives were the need to identify physical activity research newspaper reports immediately upon publication, and to collect information about the context of their production. We conclude with a discussion of the importance of methodological transparency for understanding the "meaning systems" imposed in media content analysis or qualitative content analysis more generally. [6]

## **2. Researching Health and Physical Activity in the Canadian Mass Media**

We conducted a multi-layered analysis of the social construction of health and physical activity messages in the Canadian media. Spanning three years, this project analyses the formation, production, transmission and eventual consumption of news media messages about health and physical activity. During our initial stage (November 2004-April 2005), we tracked the coverage of health and physical activity reporting in newspapers and national television news media to conduct two interconnected studies. The first examines transmission using a content analysis which asks: What is reported about physical activity and health in the Canadian news media (see FAULKNER, FINLAY & ROY, in press)? The second, arising from the content found in the first, focuses on issues surrounding the production of physical activity research stories, particularly the role of sources and journalists (see FINLAY, ROY & FAULKNER, 2006). Journalists and sources were identified from the newspaper articles and asked to participate in an interview or complete a questionnaire; therefore central to the process of data collection was the initial determination of articles to include. [7]

### **2.1 Locating relevant media texts**

It was necessary to choose a method of identifying and collecting newspaper articles on physical activity and health that satisfied the objectives of our research. Two important criteria were identified: First was the timely identification and collection of all physical activity research articles. Finding these articles quickly was crucial to our ability to contact journalists and sources while the processes of production were still fresh. Second, we were interested in not only the text of these stories, but the images, the placement of the stories within the newspaper and other information that can only be found in the original newspapers. This meant it was necessary to collect newspaper stories in their original form through either original clippings or photocopies in order to understand the context of their production. [8]

### *2.1.1 Sampling*

According to JENSEN (2002), sampling for content analysis requires several steps. First is to specify which content will be sampled. Our choice of newspaper titles reflected both a) national and regional coverage which reaches the highest number of readers based on circulation figures, and b) pragmatic concerns—hard copies of the titles had to be available locally. This broad sample allowed for intensive sampling of media coverage on issues of health and physical activity, as well as providing a means of comparison between different reporting styles and contrasting news media. [9]

Second is a determination of how many editions will be examined and over which time period (JENSEN, 2002). Our sampling method was prospective rather than retrospective. Sampling prospectively means collecting materials as they are produced in order to examine the most recently published media available (DEACON, PICKERING, GOLDING & MURDOCK, 1999). A six month period was chosen because it was deemed to be a comprehensive time period for analysis in light of available time and resources. [10]

The third step is to determine what counts as a news story and how these are defined and delimited (JENSEN, 2002). At the outset we decided that stories which reported on any aspect of physical activity would be collected to provide the context within which physical activity discourse is created in the news media. This would form the basis of the content analysis. In cases where physical activity research stories were found (often involving a close reading of the article), the sources and journalists were identified and contacted (by mail and telephone respectively) and were asked to complete a questionnaire or participate in an interview about their role in the production of this particular report, and their perceptions of media reporting of health and physical activity research more generally. [11]

### *2.1.2 Search and retrieval*

With these research objectives and sampling procedures established, print indexes were disqualified due to the time-sensitive nature of our research. As noted earlier, print indexing services can take many months to a year to be produced and while they can be useful for historical research, these indexes do not enable the prospective sampling of media accounts as they unfold. Computerised indexes were also rejected because the contextual information (specifically images, placement on the page and headline size) is lost when newspaper articles are turned into searchable text-only files. [12]

To locate, in a time-sensitive manner, an original copy of the articles to document this information seemed to necessitate hardcopy hand-searching. This involves scanning each page to locate articles that discuss the relevant topic. However, the hand-search method of tracking stories is very time-consuming and the sheer bulk of media products (in our case over 50 newspapers a week) can be

overwhelming; as a result we adopted a multiple-methods approach to our newspaper search plan. [13]

In order to lessen the scope of the media to be tracked, a number of strategies can be implemented which combine different search methods. First, a daily hardcopy search of those news outlets considered to be "agenda-setters" can assist in highlighting or identifying issues which have a national scope. We conducted a daily hand-search of the three newspapers with the largest national circulation—*The Toronto Star*, *The Globe and Mail*, and *The National Post*. The remaining eight newspapers were hand-searched on a rolling basis by alternate days (four on Monday, Wednesday, Friday, Sunday and four on Tuesday, Thursday, Saturday, with the titles switching sequence each week). Second, this was backed up with a computerised keyword search of an online newspaper database. This did not replace the hardcopy searches or the collection of articles from the original newspapers, but allowed us to flag articles with health and physical activity research content so we could contact the sources and journalists in a timely manner. [14]

Using this combination of methods proved much more time efficient for sample identification. However, rarely having encountered discussions of these methodological choices in research using content analysis, we were concerned about losing the applied process of hand searching and chose to compare the hardcopy hand-search with a keyword computerised search of the *Newslink* database and *Factiva* (a more widely available newspaper computerised search engine). [15]

## 2.2 Search and retrieval comparison

Two newspapers (*The Toronto Star* and *The Globe and Mail*) were independently tracked for articles using the key words: exercise, fitness, obesity, overweight, physical activity and physical fitness for a one-month period from December 20, 2004 to January 20, 2005. The second author conducted hardcopy searches using the original newspapers to scan for relevant content. The first author conducted key word searches using the *Newslink* database and *Factiva* online index provided through the University of Toronto Library. We compared the results of the searches to gauge the level of agreement between the two methods. [16]

The electronic search indexes yielded many more hits (articles containing the key words) than the hardcopy searches. This is because the computer program will find every instance of the search terms, resulting in a high number of false positives; items identified by the computer but not relevant to our research (SOOTHILL & GROVER, 1997). As a result, all citations that were not about health and physical activity (i.e. discussions of a military "exercise", politicians throwing their "weight" around) were weeded out. [17]

Conversely, hardcopy searches can result in a number of false negatives (the missing of articles which should be included) (SOOTHILL & GROVER, 1997).

Hardcopy hand-searches do not involve reading every story thoroughly as this would be exceedingly time consuming. Instead, headlines and photographs are scanned, and the researcher then skims potentially relevant stories. It is a process where the reader can get bored or distracted, and where relevant content can be hidden in final paragraphs, advice columns, letters to the editor and other areas where the content might not be readily apparent (SOOTHILL & GROVER, 1997). [18]

For *The Toronto Star* the hand-search method resulted in 21 stories, and the online indexes each located 27 stories that fulfilled our search criteria (there was agreement between the two electronic indexes in all cases). The hardcopy search located one article that was not found using the computer keyword search and would therefore be a false negative had only a computer search been conducted. The article was reprinted from a foreign newspaper and we discovered that the computerised databases did not index articles that were reprinted from newspapers that did not grant copyright to these particular programs. While this story referenced health and physical activity, its non-Canadian focus would exclude it from further analysis.

Type of article	Headline	Clear relevance?
Weight Loss	Atkins is out, slim is always in	yes
Weight Loss	Low-fad diets key, nutritionists warn	yes
News Brief	Brief: Ontario urban sprawl	no
Business Article	Kraft food flap	no
Cooking feature	Boost your grain power	no
Advice Column	Counselling would help marriage and sex life	no
Popular Culture	Spotlight	no
Popular Culture	Looking for a second helping of fame	no
Humour Column	Going whole hog	no

Table 1: Articles not located using a hardcopy hand-search method for the *Toronto Star* [19]

The online keyword search yielded nine articles that were not located by hardcopy searching. A summary of these can be found in Table 1. In four instances a keyword was mentioned only once or very briefly in the articles and did not have obvious health content. The first two articles in the table were about weight loss and fit our search criteria. They were likely missed during the hardcopy hand-search due to human error. [20]

*The Globe and Mail* search yielded similar results. The hardcopy hand-search found 28 articles and the electronic keyword search resulted in 43 articles. Only one article found during the hardcopy search was missed by the computerised

search. It was an article announcing a midnight run and yoga session to celebrate New Year's in Toronto. *The Globe and Mail* prints more than one edition and this was local content that may not have been indexed. It is important to account for possible regional editions printed by one newspaper and ascertain which edition is catalogued by the chosen index.

Type of article	Headline	Clear relevance?
Focus Section	Exercise: How to take your outdoor sport inside for winter	yes
Health	I can't believe I just ate the whole figgy pudding	yes
News	Alberta to study two-tiered health care	no
Travel	Promising journeys	no
Travel	Dangling the carrot of travel	no
Letter to Editor	Celebex Too?	no
Business	Kirstie Alley signs up to front weight-loss ads	no
Facts & Arguments	Social Studies: A Daily Miscellany of Information	no
Facts & Arguments	Social Studies: A Daily Miscellany of Information	no
Facts & Arguments	Social Studies: A Daily Miscellany of Information	no
Cooking	Bread is back for 2005	no
Letter to Editor	Revolutionary idea?	no
Popular Culture	Fat actress: three clicks past Rubenesque	no
News	Inside City Hall	no
Column	You say you want a resolution	no
Health	Quitting, starting, quitting, starting: ending the cycle	no

Table 2: Articles not located using a hardcopy hand-search method for *The Globe and Mail* [21]

There were 16 articles found with the keyword search of the online indexes not found by hand-searching. A summary of these articles can be found in Table 2. Eight articles contained a one-line or very brief mention of the keywords, and in all cases it would be difficult to find the keywords without reading the entire text of the articles. However there were two articles missed using the hand-search method due to human error. In both cases, the computerised keyword search yielded more articles than the hardcopy hand-searches. Clearly, the electronic keyword search was more thorough for media tracking and content analysis research. [22]

### **2.3 Is more better? The implications of combining search strategies**

Using a combination of methodologies to develop a fuller sample in an efficient manner proved effective in this study, but it can also change the sample frame. While the electronic searches certainly created a large number of false positives which necessitated the pruning of articles which were not relevant to the study, it also created a pool of "fringe" articles which could be included in the study, but likely would not have been located had we relied strictly on print indexes or hardcopy searches. For example, a large number of articles were identified that mentioned physical activity once only or very briefly. These instances could be considered interesting because they may point to physical activity's relative lack of newsworthiness compared to other health topics and discussions of disease. The call to exercise regularly for good health and prevention of disease is a common message that is taken as "truth" by most in society, and it was often positioned as an "add-on message" in articles on health-related topics like diabetes, cancer, obesity and heart disease (e.g. in our sample of 1011 articles, 11.2% qualify as including only a passing reference to health and physical activity; see FAULKNER et al., in press). Discursively the mention of physical activity is important as it demonstrates how exercise is considered an imperative for good health, but also, like other messages (i.e. eat right and don't smoke) it is not really news and does not warrant the coverage afforded to high-profile diseases and medical breakthroughs. This was also substantiated by our interviews with journalists. As a result these articles were included in our final sample. [23]

The ease of using the computer to locate these articles and others which were tangential to the project (i.e. obituaries, business announcements and celebrity gossip) creates the temptation to broaden the original search criteria. Computerised keyword searches have many benefits; however they could also change the nature of qualitative research towards a focus on minutia or the privileging of quantitative measures over qualitative ones. In a quantitative content analysis these stories could be counted toward the frequency of news reports on physical activity, however it is only through a closer qualitative reading of these texts that the ways in which these stories frame physical activity can be more fully explained. Researchers need to be clear about these issues at the outset of their studies, when choosing their sampling criteria, search tools, and in reporting their findings. [24]

The use of the online keyword search proved to be a more reliable, easier and quicker way to locate newspaper articles than hardcopy hand-searches. However, we have identified limitations to this method that may necessitate the use of a mixed approach to data collection. The text-only versions provided by the databases remove articles from their original form which can carry important information about the visual and spatial configurations of newspapers (SOOTHILL & GROVER, 1997). In addition, occasionally while locating articles in the actual newspaper using the computerised search results as a guide, another article might be found. Often these articles were not contained in the computer index due to copyright restrictions. Combining the two methods allowed us to

serendipitously locate some false negatives that we otherwise would have missed. Computerised searches are not perfect and their limitations need to be recognised. [25]

### **3. Conclusions**

The purpose of this article was to step back from the usual details of method which are reported on content analysis to examine the implications of some of the decisions that are made in determining sample production. It is important for qualitative researchers working with media texts to be clear about the methods which are used in creating a sample. In addition, it is wise to conduct a reliability study at the outset to ensure that one's chosen method is reliable, provides the kinds of data required and fulfils the research objectives (see MANEY & OLIVER, 2001). [26]

These issues are also generalisable to a broader audience of qualitative researchers. Consider, for instance, the increasing use of computer programs for supporting qualitative research (e.g., NUD\*IST). Researchers using such programmes to catalogue narrative data may use it as a means to find every instance when someone refers to, for example, their head. Using the search function, a search could be conducted using keywords such as "head", "noggin" and "nut". Clearly in a large corpus of data this would throw up a number of false positives, for instance when an interviewee referred to a "head start" or a "peanut" butter sandwich. At the same time, it might miss instances where interviewees used unusual phrases or combinations of words (e.g., "the space between my ears") which may only be found, and understood, within the context of the surrounding text. Computer programs certainly aid the process of qualitative content analysis, however they cannot fully replace researchers' interpretations of texts (MAYRING, 2000). [27]

In our case, while hand searching original copies of the newspapers allowed us to examine relevant newspaper articles in their original context and take note of visual features such as images and placement on the page, it was also time-consuming and tedious work which can be less thorough and accurate than computerised key-word searches. The use of electronic search databases permits the easier and more rapid retrieval of newspaper articles, but also tends to create many false positives which must be sorted through by the researchers. Our choice of search methods had implications for the sample produced. The ease of searching and the broad range of results created the temptation to include media pieces which may be tangential or even irrelevant to the original research goals. The imposition of meaning systems on the frames surrounding content analysis may occur at very early stages in the research. Overall, it is important that qualitative researchers provide methodological transparency about the choices and decisions they make in determining their sample. [28]

## Acknowledgements

This study was supported by a research grant from the Social Sciences and Humanities Research Council of Canada.

## References

- Deacon, David; Pickering, Michael; Golding, Peter & Murdock, Graham (1999). *Researching Communications*. London: Arnold.
- Finlay, Sara-Jane; Roy, Stephannie C. & Faulkner, Guy (2006). *Translating health: The professional practices of journalists and sources in reporting on health and physical activity in the Canadian mass media*. Paper presented at the 36th Annual Popular Culture Association Conference, Atlanta Georgia, April 12-15.
- Faulkner, Guy; Finlay, Sara-Jane & Roy, Stephannie C. (in press). Get the news on physical activity research: A content analysis of physical activity research in the Canadian print media. *Journal of Physical Activity and Health*.
- Jensen, Klaus Bruhn (2002). *A handbook of media and communication research*. London: Routledge.
- Krippendorff, Klaus (2004). *Content analysis: An introduction to its methodology*. Thousand Oaks, CA: Sage Publications.
- Maney, Gregory M. & Oliver, Pamela E. (2001). Finding collective events: Sources, searches, timing. *Sociological Methods and Research*, 30(2), 131-169.
- [Mayring, Philipp](http://www.qualitative-research.net/fqs-texte/2-00/2-00mayring-e.htm) (2000). Qualitative content analysis [28 paragraphs]. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 1(2), Art. 20, <http://www.qualitative-research.net/fqs-texte/2-00/2-00mayring-e.htm> [Date of Access: October 26, 2006].
- McQuail, Denis (2005). *McQuail's Mass Communication Theory*. London: Sage Publications.
- Pearson, Jayn, & Soothill, Keith (2003). Using an old search engine: The value of *The Times Index*. *Sociology*, 37(4), 781-790.
- Soothill, Keith & Grover, Chris (1997). A note on computer searches of newspapers. *Sociology*, 31(3), 591-596.
- van Zoonen, Liesbet (1994). *Feminist media studies*. London: Sage Publications.
- Zollars, Cheryl (1994). The perils of periodical indexes: Some problems in constructing samples for content analysis and culture indicators research. *Communication Research*, 21(6), 698-716.

## Authors

Stephannie C. ROY PhD., is a Post-Doctoral Researcher in the Faculty of Physical Education and Health at the University of Toronto. Her research interests centre around the practices of health communication in the media including the processes of inception, production and reception and qualitative analyses of media representations of women's health.

Contact:

Stephannie C. Roy

Faculty of Physical Education and Health  
University of Toronto  
55 Harbord Street, Toronto, ON, M5S 2W6  
Canada

Tel.: 416-946-0262

Fax: 416-971-2118

E-mail: [stephannie.roy@utoronto.ca](mailto:stephannie.roy@utoronto.ca)

Guy FAULKNER PhD., is an Assistant Professor in the Faculty of Physical Education and Health at the University of Toronto. Guy's research interests lie primarily within the field of physical activity and psychological well-being. He is also interested in qualitative research. In particular, the use of ethnographic techniques, and conversation analysis are of interest.

Contact:

Guy Faulkner

Faculty of Physical Education and Health  
University of Toronto  
55 Harbord Street, Toronto, ON, M5S 2W6  
Canada

Tel.: 416-946-7949

Fax: 416-971-2118

E-mail: [guy.faulkner@utoronto.ca](mailto:guy.faulkner@utoronto.ca)

Sara-Jane FINLAY received her doctorate in Media Sociology from Loughborough University in the UK. She taught in Southampton and Plymouth before returning to Canada. While still teaching in media and visual culture, she is the Director, Academic Human Resources in the Office of the Vice President and Provost at the University of Toronto where she is responsible for the recruitment, integration and retention of faculty and conducts research on the faculty life cycle with an emphasis on equity and diversity.

Contact:

Sara-Jane Finlay

Office of the Vice President and Provost  
University of Toronto  
27 King's College Circle, Toronto, ON, M5S  
1A1  
Canada

Tel.: 416-978-1855

Fax: 416-978-3939

E-mail: [sarajane.finlay@utoronto.ca](mailto:sarajane.finlay@utoronto.ca)

## Citation

Roy, Stephannie C.; Faulkner, Guy & Finlay, Sara-Jane (2007). Hard or Soft Searching? Electronic Database Versus Hand Searching in Media Research [28 paragraphs]. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 8(3), Art. 20, <http://nbn-resolving.de/urn:nbn:de:0114-fqs0703204>.