Computational Methods for Analyzing Health News Coverage

Delano J. McFarlane

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
under the Executive Committee
of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2011

ABSTRACT

Computational Methods for Analyzing Health News Coverage

Delano J. McFarlane

Researchers that investigate the media's coverage of health have historically relied on keyword searches to retrieve relevant health news coverage, and manual content analysis methods to categorize and score health news text. These methods are problematic. Manual content analysis methods are labor intensive, time consuming, and inherently subjective because they rely on human coders to review, score, and annotate content. Retrieving relevant health news coverage using keywords can be challenging because manually defining an optimal keyword query, especially for complex health topics and media analysis concepts, can be very difficult, and the optimal query may vary based on when the news was published, the type of news published, and the target audience of the news coverage.

This dissertation research investigated computational methods that can assist health news investigators by facilitating these tasks. The first step was to identify the research methods currently used by investigators, and the research questions and health topics researchers tend to investigate. To capture this information an extensive literature review of health news analyses was performed. No literature review of this type and scope could be found in the research literature. This review confirmed that researchers overwhelmingly rely on manual content analysis methods to analyze the text of health news coverage, and on the use of keyword searching to identify relevant health news articles.

To investigate the use of computational methods for facilitating these tasks, classifiers that categorize health news on relevance to the topic of obesity, and on their news framing were developed and evaluated. The obesity news classifier developed for this dissertation outperformed alternative methods, including searching based on keyword appearance. Classifying on the framing of health news proved to be a more difficult task. The news framing classifiers performed well, but the results suggest that the underlying features of

health news coverage that contribute to the framing of health news are a richer and more useful source of framing information rather than binary news framing classifications.

The third step in this dissertation was to use the findings of the literature review and the classifier studies to design the SalientHealthNews system. The purpose of SalientHealthNews is to facilitate the use of computational and data mining techniques for health news investigation, hypothesis testing, and hypothesis generation. To illustrate the use of SalientHealthNews' features and algorithms, it was used to generate preliminary data for a study investigating how framing features vary in health and obesity news coverage that discusses populations with health disparities.

This research contributes to the study of the media's coverage of health by providing a detailed description of how health news is studied and what health news topics are investigated, then by demonstrating that certain tasks performed in health news analyses can be facilitated by computational methods, and lastly by describing the design of a system that will facilitate the use of computational and data mining techniques for the study of health news. These contributions should further the study of health news by expanding the methods available to health news analysis researchers. This will lead to researchers being better equipped to accurately and consistently evaluate the media's coverage of health. Knowledge of the quality of health news coverage should in turn lead to better informed health journalists, healthcare providers, and healthcare consumers, ultimately improving individual and public health.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF EQUATIONS

ACKNOWLEDGEMENTS

The completion of this dissertation was made possible by the guidance, support and friendship of numerous people. First I must acknowledge my advisor, Dr. Rita Kukafka, for supporting me in my research and challenging me to make my work both relevant and robust. I also thank my other committee members, Drs. Stephen Johnson, Noemie Elhadad, Jessie Gruman and Jessica Ancker, for all the insight and assistance they have so generously provided. Thanks also goes to Marguerite Holloway who also provided advise and support throughout my research.

I would also like to thank Dr. Dan Stein, Dr. Jesse Wrenn, Karthik Natarajan, and Dr. Samuel Brody for their support in helping me refine and clarify my ideas and methods, and Dr. George Hripcsak for his thoughts on my overall research. Dr. Richard Sloan also deserves special thanks for his support during my time at Columbia and for the many years prior. The personal and professional relationship that we have developed over the years is one I value deeply.

Thank you to all those that generously volunteered their time during the development of the gold standards used in the evaluations in this research. Your work was instrumental and greatly appreciated.

This research was made possible by NLM training grant LM007079-15, and NIH NCMHD Dissertation Research Award R36.

And lastly I must acknowledge my family and friends, in particular my parents Yvonne McFarlane and Vanril McFarlane who have been unfailing in their love and support. Thanks and much love.

Chapter 1. INTRODUCTION

"'Health News', through cupidity or ignorance of its sponsors, may be quite as injurious to public welfare as the practice of the healing art by the incompetent or unscrupulous." - Unknown author; 1926[1]

"Few intelligent observers will question the statement that the press influences humanity en masse in health matters far more than do doctors and other health agencies" - Unknown author; 1926[1]

1.1. Problem statement and significance

Research shows that the public's knowledge and attitudes towards health issues are influenced by media coverage, which can in turn influence individual and public health[2-4]. Unfortunately studies continue to find that health news coverage is often biased, inaccurate or incomplete[5-11]. To identify and combat shortcomings in health news it is essential that researchers monitor and evaluate the media's coverage of health, and communicate their findings to journalists, health officials and health communications professionals[12].

Health news investigators have historically relied on the manual reading, and scoring of news text[13-15], and on the use of simple keyword queries for the coding and retrieval of relevant health news coverage[14,16,17]. These methods are problematic. An accurate, manually created keyword query can be hard to define for conceptually complex topics, and may vary based on numerous factors including news source, type of news publication, date of publication, and the news publisher's target audience. Manual review and coding of news articles relies on content analysis methods that are very time consuming, and labor intensive. In addition the inherent subjectivity introduced by using of human coders must be mitigated with consensus building methods (e.g. Delphi method[18]) that are themselves both time and labor intensive.

Lastly, advances in information technology have drastically changed how news is generated, distributed and consumed. Traditional sources of health news now compete with health blogs[19], health websites, and other online sources of health information such as Wikipedia[20]. The Internet has transformed the landscape of health news into a data rich environment where there is vastly more health news content produced today than every before. The methods used by researchers to evaluate health news coverage should combine the theories and concepts traditionally used in health news analyses with computational and data mining techniques that can scale to meet the challenges and diversity of the health news landscape of today.

1.2. Not just about text classification

There are numerous computational methods from the research fields of information retrieval (IR), natural language processing (NLP) and machine learning capable of automatically grouping, characterizing and classifying news articles in more robust ways than keyword filtering, and scoring news articles on various dimensions ranging from word use and syntax, to affective emotions. However, there are aspects of the classification, retrieval, and scoring tasks performed in health news investigations that differentiate them from basic document classification, analysis and retrieval problems.

Typical machine learning systems require training data sets that range in size from hundreds to many thousands of cases. The cost associated with manually generating correct classifications of news content for health news investigations can be high, requiring that human coders be recruited and trained, then having them read and annotate individual health news articles, and often requiring that they reach consensus with other coders. Additionally, many health news studies evaluate much smaller corpora than would be needed to train a typical machine learning classifier. Since the time and resources required to generate a typical training dataset may by too burdensome to many researchers in this field, any system developed to facilitate health news investigations must demonstrate acceptable levels of performance using fewer training cases than are typically required, or minimize the work required to attain sufficient training cases.

Another factor is that news coverage relevant to a health topic may vary across a corpus based on many factors including what other health issues are discussed with the topic of interest, the publication date of the news, and the level of relevance required. To account for these factors multiple classification features need to be used together to produce accurate classifications, or features must be reported independently so that investigators can draw their own conclusions.

Another issue is that the health news topics that researchers are interested in vary considerably. Many researchers are concerned with broad, long-lived health issues that journalists over time have developed well-defined ways of discussing (refer to the media template discussion in Section 3.2 on page 22). However, researchers are also interested in novel, short lived, event driven health news stories, for which there is no established template for discussion. Still other researchers are not as interested in *what* is being discussed (e.g. a well known disease) as they are in *how* it is being discussed (refer to the framing media effect discussion in Section 3.1 on page 20). Classification and scoring algorithms developed to facilitate the analysis of health news coverage must be flexible enough to handle the variety of classification and scoring tasks performed by health news investigators.

Lastly, investigators need a way of approaching the study of health news coverage as both a content analysis and as a data mining exercise. This means that the content analysis methods and concepts currently favored by investigators must be merged with concepts from the computational sciences in a coherent and cohesive manor.

1.3. Obesity news, framing, populations, and health disparities

Unique, relevant, meaningful challenges were selected to evaluate and explore the use of the tools and algorithms developed in this dissertation. The classification and scoring tasks explored were, (1) the retrieval and classification of health news relevant to the topic of obesity, (2) classifying health news based on the manner in which the news was framed, and (3) the investigation of how health news varies when different populations are being discussed.

Obesity news

The retrieval and classification of obesity news was selected as a task for investigation because of the importance of obesity as a public health issue, and the diversity of contexts within which obesity news is discussed. Obesity is one of the leading avoidable causes of death in the world[21]. The media's coverage of obesity has been of interest to researchers since at least 1985 when a National Institutes of Health panel determined obesity to be a major health concern[22,23]. That interest still exists today as evidenced by the regularity with which health news researchers study news coverage of obesity (refer to the literature review results in Chapter 3 on page 34).

The classification of obesity news was also selected because of the variety of concepts often associated with obesity, and the various ways in which obesity is discussed in the news (e.g. personal responsibility, treatment options, prevention, environment risk factor for other health issues) [23-25]. This diversity of news coverage makes obesity news retrieval and classification tasks relevant, challenging, and potentially generalizable to other health news topics.

Media effects analysis

The analysis of news coverage can be organized into 3 major types, agenda setting studies, framing studies and informing studies. In **agenda setting** studies researchers attempt to identify news stories that the media imply are important based on the volume of news coverage for that story. For example, if researchers find that over some period of time the media publish more obesity news compared to news of other health issues, then researchers can infer that over that period of time people who follow the news (and potentially people who don't) will think obesity is a more newsworthy, and possibly a more important topic than other health issues. **Framing** studies explore *how* the media report on a news story. For example researchers may be interested in knowing if news coverage about obesity tends to emphasize personal responsibility, over other aspects of the disease such as biological and genetic factors. **Informing** studies investigate the accuracy, completeness and presentation of the information present in media coverage. For example researchers may be interested in if news coverage of medical research includes information such as where the research was conducted, specifics about the populations used in the research, and what the immediate impact of the research might be.

In this dissertation, computational methods that facilitate tasks performed during the analysis of news coverage, in particular tasks required for agenda setting and framing analyses, were studied. Agenda setting was selected because the primary task involved in agenda setting studies, categorizing news coverage by topic, is applicable to all health news evaluations. This is because the first task in evaluating media coverage is typically the acquisition of a corpus of news content that is relevant to the research questions being investigated.

The facilitation of classification and evaluation tasks performed during framing studies was investigated for 3 reasons; (1) classification and scoring of health news frames presents a challenging and complementary task to categorizing health news on broad health topics, (2) framing is frequently studied by health news investigators, (refer to the literature review results in Chapter 3 on page 34), and (3) an understanding of the organizing ideas that exist in health news coverage is critical to knowing what messages the public is receiving concerning newsworthy health information.

This dissertation does not specifically investigate the development and evaluation of computational methods that would facilitate tasks intrinsic to informing studies. The facilitation of tasks fundamental to informing studies presents different challenges than the classification tasks investigated in this research. Future research that applies automated information extraction, and summarization techniques should be investigated for facilitating informing studies.

For a more detailed description of media studies research refer to Section 3.1 on page 16.

Populations and health disparities

To demonstrate the methods and algorithms investigated in this dissertation, the framing of health and obesity news that discusses specific populations (e.g. children, women, seniors, African Americans, Latinos) was explored. Numerous disparities in the health of specific populations have been documented[26,27]. Health news researchers regularly perform studies to see if these disparities are reflected, influenced or reinforced by news coverage[28-31]. In these studies investigators must identify news coverage that refers to specific populations, categorize or score those news articles based on their framings, and analyze the scoring results to identify significant findings. The importance of exploring health disparities as they are manifested in health news coverage, along with the various classification, scoring and evaluation tasks involved in this investigation made it appealing as a way to illustrate the methods and concepts explored in this dissertation.

1.4. Research questions, aims and hypotheses

This dissertation has 3 stages. The first is a literature review of health news analyses. This review was conducted to identify the methods that health news researchers most frequently use, and the research questions and health news topics researchers frequently investigate. The second stage is a study that applied computational methods to the basic classification, information retrieval, and scoring tasks typically performed in health news studies. The last stage is the design and demonstrated use of SalientHealthNews, a system that brings together the concepts and findings of the first two studies into a health news analysis workbench that incorporates algorithms and tools for the retrieval, annotation, classification and analysis tasks performed during health news investigations. The aims, hypotheses and research questions for each of these studies are listed on the next 2 pages.

Table 1. Research aims, questions and hypothesis

Research aims	Research Questions	Hypotheses
Health news analysis literature review: Identify the methods that are typically used, and the research questions and health topics that are frequently investigated by health news researchers (Chapter 3).	1) What methods do investigators favor when analyzing and retrieving relevant health news coverage? 2) What health news issues and research questions do investigators frequently explore? 3) Do researchers continue to find that health news coverage is deficient?	1) Researchers favor (a) keyword querying for retrieving relevant news coverage, (b) manual content analysis for coding and scoring, and (c) basic statistical significance tests such as Student's *t*-test and the chi-square test. 2) Researchers most frequently perform agenda setting studies over framing studies, and framing studies over informing studies. 3) Health news coverage continues to have considerable shortcomings in quality.

Table 1. Research aims, questions and hypothesis (continued)

Research aims	Research Questions	Hypotheses
Classifier development and evaluation: Develop and evaluate an obesity news classifier and health news framing classifiers. (Chapters 6 and 7)	1) What level of performance can be achieved for classifying health news on relevance to obesity and on media framing using semi-supervised classification methods? 2) What classification features perform best for these tasks?	1) There are patterns in the words and language of health news text that can be used to automatically classify health news by topic and framing. 2) Classifying based on a news article's framing is a more difficult task than classifying based on broad health topics such as obesity. 3) Acceptable classifier performance can be achieved with limited training data.
SalientHealthNews: Describe and illustrate the use of SalientHealthNews, a system that incorporates a suite of annotation, computation, content analysis and statistical tools to facilitate the investigation of health news coverage (Chapter 8).	1) How can the literature review and classification findings inform the design of a health news analysis and data mining system? 2) Are there other methods that can be used to facilitate tasks performed while evaluating health news coverage?	

1.5. Dissertation Outline

In the next chapter I provide an overview of the various research domains that contributed to this dissertation, and the process I used to explore those domains. In Chapter 3 I present background research that describes the theoretical framework for the evaluation of health news. In Chapter 4 I describe the methods and findings of the health news analysis literature review conducted for this dissertation. In Chapter 5 I describe various computational methods that were used to develop the classifiers evaluated later in this dissertation. Chapter 6 details a preliminary study conducted to test the feasibility of *n*-gram probabilistic language modeling for differentiating obesity news coverage from more general news text. In Chapter 7 I describe the methods used to implement and evaluate an obesity news classifier, and multiple health news framing classifiers. Results and a discussion of the findings of the classification study are presented in Chapter 8. Chapter 9 describes the design of the SalientHealthNews system and presents preliminary data generated by SalientHealthNews algorithms. Chapter 10 concludes by discussing the significant findings and the major contributions of the dissertation research.

1.6. Summary

In summary, this dissertation investigates how the researchers evaluate health news and explores the use of computational methods in the annotation, scoring, and analysis of health news text. The successful use of these methods will better equip investigators to continue the important work of evaluating media coverage of health.

Chapter 2. BIOMEDICAL INFORMATICS - A DOMAIN OF DOMAINS

The diversity of knowledge required to pursue research in biomedical informatics can present challenges. How does a researcher coherently navigate the various research domains that contribute to biomedical informatics? How does a researcher gain the breadth and depth of knowledge needed to study their research goals? This chapter briefly discusses these questions in the context of this dissertation's research. In the first section of this chapter the various scientific domains that contributed to this dissertation are discussed. In the second section the process used to iteratively explore these domains and refine this dissertation's research questions is described.

2.1. Domains contributing to this dissertation

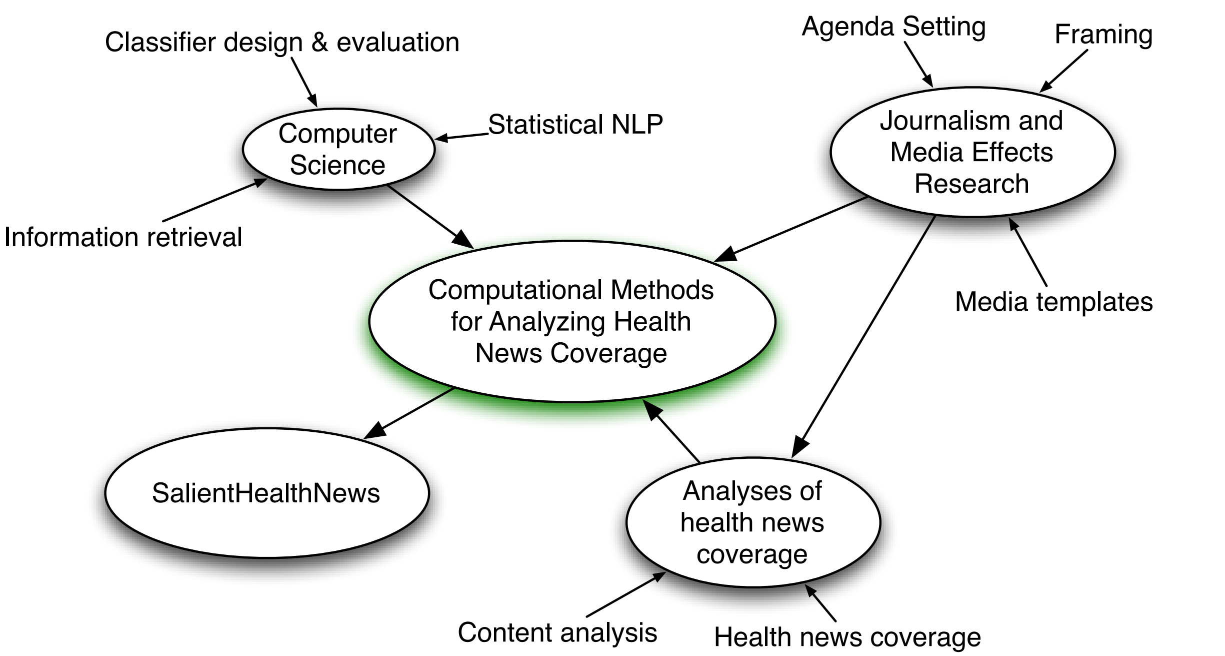


Figure 1. Mind map of research domains contributing to this dissertation

This dissertation is an example of how biomedical informatics research can span numerous, diverse research domains. For this dissertation, information retrieval, classifier design, and natural language processing concepts and methods were borrowed from the field of computer science. Content analysis methods that define how to collect and validate human coded data were borrowed from social science research. Theoretical concepts that describe the construction and analysis of news were taken from journalism and media effects research. The methods and interests of health news analysis researchers were discovered from an extensive review of health news research literature. And throughout the dissertation research, numerous statistical methods from biostatistics, social science research, information theory, and computer science were used.

These various scientific domains interacted and contributed in numerous ways, illustrated in Figure 1, to inspire and support the research described in this dissertation. How were these various scientific domains explored, and how was the knowledge gained used to refine the dissertation research plan?

2.2. Iterative research plan development and knowledge acquisition

A highly iterative process was used to amass the knowledge required for this research. Critical to this process were the repeated exploration of the various contributing scientific domains, and the iterative refinement of research questions and goals. The first step in this process was to define some initial research questions for investigation, these were *"how do researchers evaluate health news coverage?"*, and *"can the study of the media's coverage of health be facilitated by computational methods?"* . Next I collected and read numerous health news analysis research articles. These research articles were retrieved using an informal, exploratory literature search of the scientific literature databases PubMed and Google Scholar using keywords such as *news*, *quantitative analysis*, *evaluation*, *agenda setting*, and *health*. The citations of referenced in the bibliographies of relevant studies were also reviewed. This strategy found numerous health news analyses that evaluated a broad range of health news topics. By reading a random and diverse collection of these studies I learned the basic concepts and methodologies that reoccur in analyses of health news coverage. Additionally, a number of research articles stood out as good templates for health news analysis research. In particular Collins' study of news coverage of the 2002 Canadian healthcare reform debate was exemplary[32]. Collins provided a detailed and comprehensive description of the theoretical framework that defines health news analysis research, directly associated that theoretical framework to her research questions, and performed a methodologically robust study.

Collins' discussion of the theoretical underpinnings of health news research led directly to the work of the news media researcher Iyengar. Iyengar's work is seminal in the field of political media analysis in describing the ways that media coverage can influence the public's opinion of newsworthy topics and events[33,34]. Numerous health news researchers, including Collins, have benefited greatly by assimilating his work into their research.

Specific tasks relevant to health news analysis that might be facilitated by computational methods were then identified. These tasks were based on the health news analysis methods described in research articles such as the Collins paper, and on the theoretical framework of news analysis developed by researchers such as Iyengar. The two tasks ultimately selected for study in this dissertation were (1) to identify health news coverage relevant to a specific health topic, and (2) to categorize health news coverage based on its news framing. Broadly speaking, these tasks are variants of the computer science problems of information retrieval (IR) and document classification. For an expansive description of information retrieval and classification concepts and methods I referred to *Pattern Recognition*, by Theodoridis[35]. *Pattern Recognition*, along with other related computer science research and reference materials, suggested numerous generic classification and information retrieval methods that facilitate the analysis and classification of text. But as discussed in Section 1.2 on page 2, the nuances of the health news research domain necessitate that any computational methods used must be selected for and tailored to the specific requirements of health news research. To select the most promising and appropriate information retrieval, and classification methods, numerous feasibility tests were performed. One example of these tests, an evaluation of the feasibility of the use of *n*-gram language models for identifying health news, is described in detail in Chapter 6.

Based on these texts certain computational methods were selected for further investigation (e.g. statistical *n*-gram language modeling), while other methods were set aside for future study. These decisions led new changes in the dissertation research questions and overall research plan, which then led back to a more in depth exploration of the health news analysis literature, this time targeting studies that categorized health news.

Many of these studies use statistical and analytical methods such as the comparison of descriptive frequencies (see Section 4.3 page 31) and statistical significance testing, to analyze the categorizations of human coders. These types of studies were frequently referred to as quantitative content analyses[36-38]. To fully understand the content analysis methods used in these studies I turned to Krippendorff's *Content analysis: An introduction to its methodology*[39]. This book thoroughly described the role and use of content analysis methods to manually categorize and score content.

With this additional knowledge it was time again to reexamine and refine the research questions and overall research plan.

This process of refinement of research plan followed by exploration of scientific knowledge followed again by research plan refinement continued until the depth and breadth of knowledge encompassed within the dissertation research plan was sufficient to produce a coherent, comprehensive and relevant array of research questions, while simultaneously providing sufficient knowledge of relevant methods, theories and concepts.

2.3. Summary

This chapter gave a broad overview of the various research domains that contributed to this dissertation. Examples of noteworthy research that helped in the development of this dissertation were give, including the media studies research of Iyengar[33,34], a health news analysis article discussing news coverage of Canadian healthcare reform by

Collins[32], and *Pattern Recognition*, a book on classification, data mining and information retrieval, written by Theodoridis[35]. In the next chapter background on media effects research and media templates will be discussed. These two concepts provide a theoretical framework for the analysis of health news.

Chapter 3. BACKGROUND: MEDIA EFFECTS AND MEDIA TEMPLATES

"Why does a dog wag its tail? Because a dog is smarter than its tail. If the tail were smarter, the tail would wag the dog." – Wag the Dog (1996)

In this chapter I discuss the two theoretical concepts, media effects and media templates, that help define the analysis of health news coverage. Section 3.1 discusses the agenda setting, informing, and framing media effects researchers have defined as the mechanisms by which the news media can influence public knowledge and perception of newsworthy topics and events. Section 3.2 describes media templates, the recurring patterns that exist in news coverage, used by journalists to structure and simplify the creation news coverage, and expected by the public to make more efficient the assimilation of news content and information.

3.1. Media Effects

Research studies that evaluate news typically estimate the features of news coverage that may influence the public's knowledge and opinion of newsworthy topics and events. Political science and communications scholars such as Iyengar, Scheufele and Entman group these features and the influences they may have into agenda setting, informing, and framing media effects[32,33].

Agenda Setting

Agenda setting is the media's ability to influence the perceived importance of an event or topic based on the volume of news coverage discussing that event or topic. Iyengar wrote that *"by covering some issues and ignoring others the media set the public agenda – they influence what people view as important issues"* [33]. Agenda setting is one of the most commonly cited media effects[32], and is the easiest to explain and measure. In agenda setting studies news articles are categorized based on a prominent topic discussed in the news coverage. In health news studies these topics include health interventions such as medication and surgery, unhealthy behaviors and risk factors such as smoking, obesity and drug use, and diseases such as diabetes and cancer[10,38,40-45]. Once news coverage has been categorized, researchers look for trends in the volume of news coverage. Examples of the types of trends investigated include how the volume of news coverage varies over time[42,46], how it differs based on location of publication[47,48], and how it compares to the volume of coverage received by other topics[49].

Researchers are also interested in how the volume of news coverage of time sensitive events tracks with the real-time evolution of those events. Examples of studies of this type include the evaluation of news coverage of disease outbreaks, natural disasters, scientific discoveries, food recalls and health promotion campaigns [10,37,48,50,51]. One example of this type of agenda setting study is Barnes' investigation of news coverage of Hurricane Katrina that was published during the hurricane's immediate aftermath[37]. Another example is Anderson's investigation of how trends in the volume of regional news reporting of bedbug infestations in the US might describe the spread of infestations nationwide[48]. Both of these studies (and other studies like them) evaluated the volume of news coverage received by their respective events as those events were taking place, and both studies linked that volume to the evolution of the events as they occurred.

Figure 2 is an example of results from an agenda setting study. This study chronicled the fluctuations in the number of tobacco related news stories published over two and a half years. In this study researchers linked peaks in the number of published tobacco news stories to the real world events that led to the dramatic increases in news coverage[52].

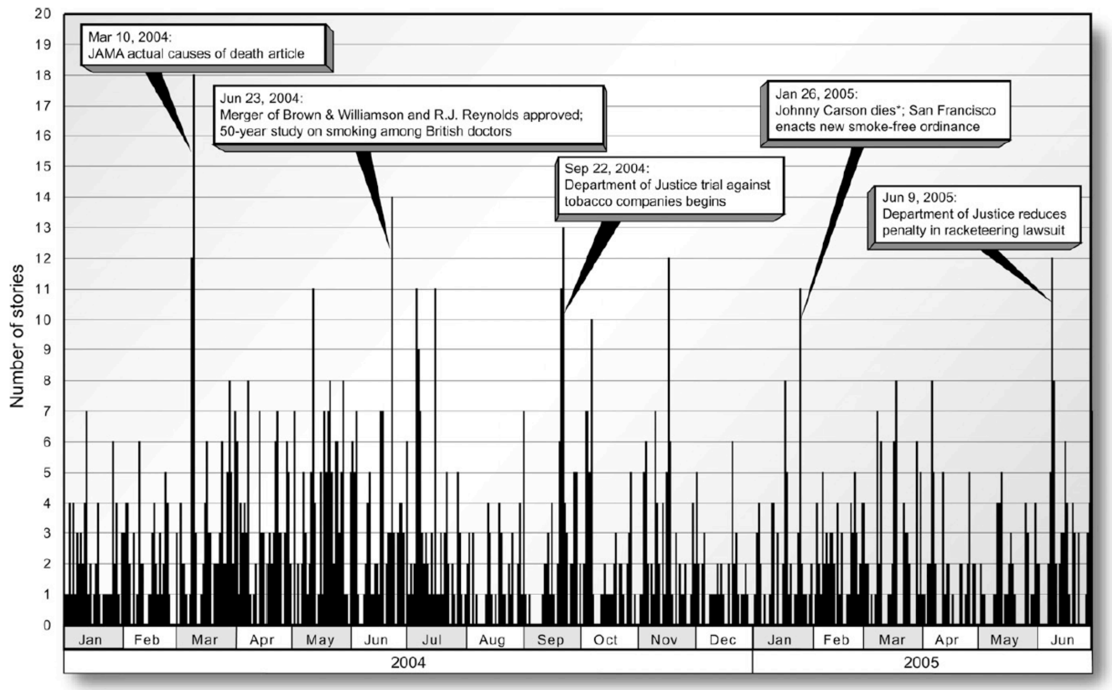


Figure 1. Newspaper and news wire service stories on tobacco, January 2004–June 2005. *Johnny Carson was a long-time smoker. News stories about his death prominently mentioned his tobacco use and that his cause of death was tobacco related.

Figure 2. Agenda setting example

<u>Informing</u>

Informing studies investigate the type and quality of information in news coverage. The types of information that researchers look for in these studies emerge *a posteriori* from a review of the news coverage relevant to the researchers' aims, or are pre-defined *a priori* using external sources of information such as health communications literature and research publications [32,53,54]. Examples of types of information examined in informing studies include the sources of information cited in news coverage, the presence or absence of detailed mortality and morbidity data, detailed research results, disease prevention and treatment options, and the risks and benefits of new treatments [55-59].

Once researchers have identified the types of information they are interested in studying, they retrieve a corpus of relevant news articles and annotate that corpus based on the appearance and quality of those information elements. Researchers then look for significant trends in content, quality, completeness and accuracy. Trends may be searched for over time, based on a news publisher's target audience, or by health news topic. Health news coverage can also be compared either to external data sources[56], or the opinions of health professionals[59].

Figure 3 is an example of results from an informing study. This study investigated how the types of information in news coverage of suicide cases differed between the years 2000 and 2006[60]. Informational elements investigated in this study included the appropriateness of the location of the news coverage, the presence of a description of the suicide method, the presence of photographs depicting the suicide scene, and the inclusion of information on suicide help services.

Table 3. Quality of reporting on individual dimensions, by year

		2000/01	2006/07
Does the item have any examples of inappropriate language? (2000/2001 *n* = 415; 2006/2007 *n* = 347)	Yes	41.7%	6.1%
	No	58.3%	93.9%
Is the item inappropriately located? (2000/2001 *n* = 415; 2006/2007 *n* = 205)	Yes	16.9%	22.9%
	No	83.1%	77.1%
Is the word "suicide" used in the headline? (2000/2001 *n* = 122; 2006/2007 *n* = 198)	Yes	29.5%	21.2%
	No	70.5%	78.8%
Is a photograph, a diagram or footage depicting the suicide scene, precise location or method used with the item? (2000/2001 *n* = 96; 2006/2007 *n* = 197)	Yes	13.5%	4.1%
	No	86.5%	95.9%
Is there a detailed discussion of the method used? (2000/2001 *n* = 232; 2006/2007 *n* = 264)	Yes	49.6%	14.0%
	No	50.4%	86.0%
Is there reference to the fact that the person who died by suicide was a celebrity? (2000/2001 *n* = 34; 2006/2007 *n* = 204)	Yes	91.2%	13.7%
	No	8.8%	86.3%
Is suicide portrayed as "merely a social phenomenon" as opposed to "being related to mental disorder"? (2000/2001 *n* = 302; 2006/2007 *n* = 216)	Yes	47.4%	23.6%
	No	52.6%	76.4%
Does the item provide information on help services? (2000/2001 *n* = 415; 2006/2007 *n* = 334)	Yes	6.5%	17.7%
	No	93.5%	82.3%
Are the bereaved interviewed? (2000/2001 *n* = 183; 2006/2007 *n* = 213)	Yes	18.0%	15.0%
	No	82.0%	85.0%

Figure 3. Informing study example

Framing

Media framing studies examine the themes and context in which newsworthy events and issues are discussed[32]. Prominent media framing researchers such as Entman and Scheufele have argued that media framing is inconsistently defined in the research literature. Entman specifically described media framing as a *"scattered conceptualization"*[61], while Scheufele argued that media framing definitions suffer from a *"theoretical and empirical vagueness"*[62]. D'Angelo went a step further arguing that not only is there no single dominant definition of news framing, but that there ought not be one. Instead, D'Angelo insisted that news framing has benefited from the constant evolving definitions and paradigms asserted by investigators of news coverage in response to their needs and the evolving nature of news coverage[63]. However, while D'Angelo makes strong points in favor of a broad, loose definition of news framing, a more specific, usable definition is required for research purposes. Two media studies researchers that provide more specific, usable definitions are Iyengar and Listerman.

Iyengar divided media framing into two types, episodic frames and thematic frames[33]. Episodic frames describe specific events and transient occurrences that have a defined and limited lifespan. Examples include disease outbreaks and natural disasters. Thematic frames are longer-lived and endemic to media coverage. They include news coverage of chronic diseases and other health topics that have had an extended "newsworthiness" lifespan, for example cancer, obesity and smoking.

Listerman divides media framing into structural and argument frames. He based his definition of framing on a definition provided by another media researcher, Gamson, who wrote that structural frames are *"a central organizing idea or story line that provides meaning to an unfolding strip of events"* [64]. Listerman added that structural frames should be *"pre-defined and based on theory, earlier research, and pre-analyses of the media text"* [45].

Argument frames, Listerman says, are the features of a news article that help promote a specific opinion or point of view. This is the only media effect that specifically measures features of news coverage that are directly and overtly intended to persuade the news audience to a specific point of view. Iyengar referred to this as the persuasion media effect[33].

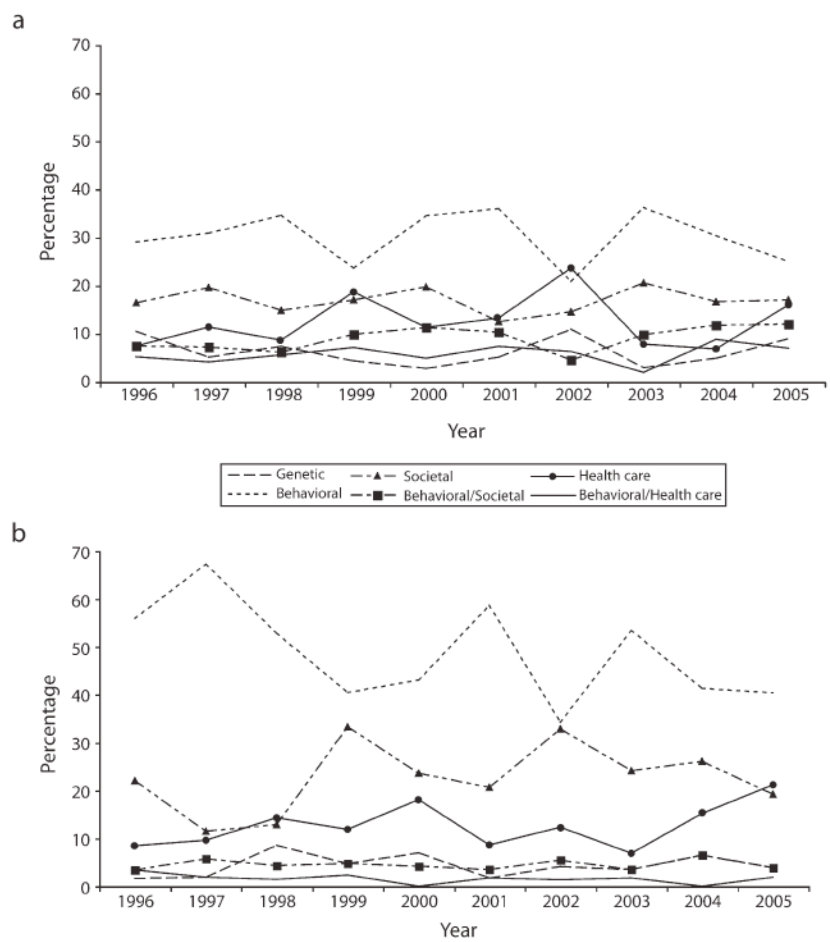
Entman described persuasion/argument framing as manifesting itself when authors:

> "… *select some aspects of a perceived reality and make them more salient in a communicating text, in such a way as to promote a particular problem definition, causal interpretation, moral evaluation, and/or treatment recommendation for the item described."* [61]

It is important to note that Iyengar and Listerman are not in disagreement on how they conceptualize media framing. Listerman's structural frames subsume both of Iyengar's episodic and thematic framing definitions, and Listerman's definition of argument frames matches well with what Iyengar's persuasion media effect.

In practice, researchers investigating the framing of health news first identify the media framings that they will be investigating, these are often broad thematic categorizations that are not specially created to match the aims or research questions being studied, but are more general and can be applied to news coverage of various topics, and research questions of various types. Examples include frames for economic issues[43], personal responsibility[42,65] and lifestyle[13,15,18].

Researchers annotate a corpus of health news either by applying the framing categorizations to units of text, or scoring text based on various attributes that contribute to the news framings being studied. Finally, researchers look for trends in the framing categorizations or the framing attribute scores in ways similar to agenda setting and informing studies (e.g. trends over time, differences between health topics).



Note. For the multilevel attribution categories, only the top 2 are presented.

FIGURE 2—Percentages of all causal and solution explanations given in articles on racial/ ethnic health disparities, by year, for (a) causal explanations and (b) solution attributions: 40 US newspapers, 1996–2005.

Figure 4. Framing example

Figure 4 is an example of results from a framing study. In this study researchers investigated how the media framed the causes and possible solutions of health disparities over a period of 10 years[31].

3.2. Media templates

One of the hypotheses of this dissertation is that there are automatically identifiable patterns in the text of health news coverage. Media templates provide a theoretical basis that supports this hypothesis. Kitzinger described media templates as a rhetorical shorthand, inspired by intense public and media interest in a seminal event, which outlives the conclusion of the event and is used to report on and make sense of new newsworthy events and public debates[66]. Media templates are created when a seminal, newsworthy event occurs, but there is no precedent for how to report on the event. Following the event a public and media zeitgeist takes form that defines the varying attitudes, biases and interests in the event and the facts and context that are salient to any description of the event. Over time this coalesces into a pattern of news coverage that addresses how the public and media define and discuss the event.

Similar to Kitzinger's media templates is another construct described earlier by Gamson called media packages[64]. Gamson suggested that, *"media discourse can be conceived of as a set of interpretive packages that give meaning to an issue"*. Gamson described how media packages manifest themselves by stating that *"a package offers a number of different condensing symbols that suggest the core frame and positions in shorthand, making it possible to display the package as a whole with a deft metaphor, catch-phrase, or other symbolic device*".

To survive and be reused media templates (or packages) must evolve, in response to changes in available information regarding real world events, due to changes in public attitude, knowledge and opinion, due to the direct efforts of individuals and organizations that wish to sway the public's feelings regarding newsworthy events and topics, or due to changes in journalistic practices. Gamson explains that packages *"succeed in media discourse through a combination of cultural resonances, sponsor activities and a successful fit with media norms and practices*", and that for packages to perpetuate themselves they must continue to construct "*meaning over time*", by incorporating news events into their structure.

To identify media templates that exist specifically in biomedical news stories, Hodgetts interviewed health journalists and asked how health news is constructed and how stories are selected[67]. The journalists said that among the considerations weighed when choosing and reporting on a health news topic are sources, the target audience, professional journalistic norms and institutional policies. The journalists also said that biomedical media templates are often simplified by having a clear plotline that presents understandable causes, consequences and solutions. All of these things influence the order with which information is presented, what information is presented and the language used to communicate that information.

3.3. Conclusion

In this chapter I described media effects, and media templates, two of the theoretical concepts that lay the groundwork for this dissertation. The basic premise of media effect theory is that media coverage impacts the public's perception of what topics are important, and provides the public with a framework for thinking about relevant news topics. Media templates describe how news stories evolve and why news coverage often exhibits a regular structure that reoccurs over time. Gamson wrote that a media template (or package) manifests itself by using "*a deft metaphor, catch-phrase, or other symbolic device*". These must ultimately reduce to the words and phrases that appear in news coverage, words and phrases that I hypothesize can be automatically recognized by computational methods.

In the next chapter I discuss a literature review of health news analyses that details the methods typically used by researchers, and the research interests often investigated by researchers.

Chapter 4. HEALTH NEWS ANALYSES: LITERATURE REVIEW

This chapter presents a literature review investigating studies of health news coverage. This review is important because the primary user of the methods and tools developed, evaluated, and used in this dissertation is the investigator of health news coverage. To ensure that this dissertation's research is relevant to them it is important to identify how they currently perform this research, and what their research interests are.

Also worth noting is that no other study of this type and scope could be found in the research literature. Only one review of studies of the media's coverage of health was found[68]. That study focused exclusively on examining research addressing the media's coverage of heart disease and gender, and is therefore different from the broader aims of this review.

In the first section of this chapter I will reiterate the research questions and hypotheses for this literature review. Section 3.2 contains a description of the methods used to retrieve relevant research and code that research for this study. Then in Section 4.3 I describe the findings of the literature review.

4.1. Research questions and hypotheses

The aims of this study were to identify what health news researchers are most interested in studying, how they typically perform their studies, and what their overall findings say about the quality of health news coverage. Three hypotheses were also defined, and emerged from the theoretical work discussed in the previous chapter, as well as an informal review of a sample of health news studies. The research questions and hypotheses for this review are listed below.

Research questions

RQ 1. What methods do investigators rely on to analyze and retrieve relevant health news coverage?

RQ 2. What health news issues are researchers interested in?

RQ 3. Do researchers continue to find that the quality of health news coverage is poor?

Hypotheses

H1. Researchers most frequently use (a) keyword querying for retrieving relevant news coverage, (b) manual content analysis for coding and scoring, and (c) basic statistical significance tests such as Student's *t*-test (for comparing average scores) and the chi-square test (for comparing categorical frequencies).

H2. Researchers most frequently perform agenda setting studies over framing studies, and framing studies over informing studies.

H3. Health news coverage continues to have shortcomings in quality.

4.2. Literature review methods

The PubMed biomedical literature database was used to identify relevant research articles for this review. PubMed is maintained by the US National Library of Medicine (NLM) at the National Institutes of Health (NIH). It contains more than 20 million citations for biomedical literature from MEDLINE, life science journals, and online books. PubMed permits users to query for research literature over various dimensions including publication type, title, abstract text, and MeSH (Medical Subject Headings)[69][*]. PubMed was the only citation database used in this review. This was done to limit the number of reviewed articles to a manageable size.

Selection criteria

An iterative process was employed to develop the query used to retrieve research articles. An initial query of ***"newspapers"[Mesh]*** returned all literature categorized with the "*newspapers*" MeSH heading. Results were examined to identify additional MeSH headings to include and exclude. This process was repeated until the results from the query were satisfactory. The final PubMed search query along with all inclusion and exclusion criteria were:

PubMed query

"newspapers"[Mesh] or ("mass media"[Mesh] or "journalism"[MeSH] AND news[title/abstract])
not Historical Article[publication Type] not Editorial[publication Type] not Comment[publication Type] not letter[publication type]
not news[publication type] not interview[publication type]

Inclusion and exclusion criteria

- Full article text available in English
- Published between January 1, 2007 and August 1, 2010
- PubMed citation data must contain the publication authors and abstract
- Research evaluates news stories directly related to health
- Research article describes in some detail the retrieval and coding methods
- Research article reports quantitative results (i.e. purely descriptive or discursive analyses[70,71] were excluded)

[*] The National Library of Medicine's controlled vocabulary thesaurus

Article analysis

Each research article was coded using the variables listed below (organized by research question):

Variables for RQ 1 (health news analysis methods)

- *Information retrieval (IR) method* – The method used by researchers to retrieve relevant news articles (e.g. manual review, and keyword filtering)
- *Coding method* – Coding/scoring method used by researchers
- *Analysis method* – Quantitative/statistical analysis conducted to analyze the coded news article data
- *Media effect* – The type of media effect analyzed in the research study (agenda setting, framing, informing)
- *Amount of news analyzed* – The volume of news analyzed

Variables for RQ 2 (researcher interests)

- *Journal* - The journal where the research article was published
- *Health topic* – The specific health news topic investigated
- *Health news category* – The type of health news being evaluated (e.g. disease, prevention, treatment, research)
- *Episodic vs. thematic* – Did the study examine news coverage linked to a specific event (*episodic*), or did it investigate news coverage of a health topic that is longer lived and has received news coverage over an extended period of time (*thematic*)
- *Country of news publication* – The country where the evaluated news coverage was published
- *Population* – Any populations specifically referred to in the research article due to the increased relevance of the health issue to that population, or because the

researchers were investigating in how news coverage varies based on the discussion of different populations

Variables for Aim 3 (news quality)

- *Quality of news* - The researchers' final assessment of the quality of the health news coverage

4.3. Results

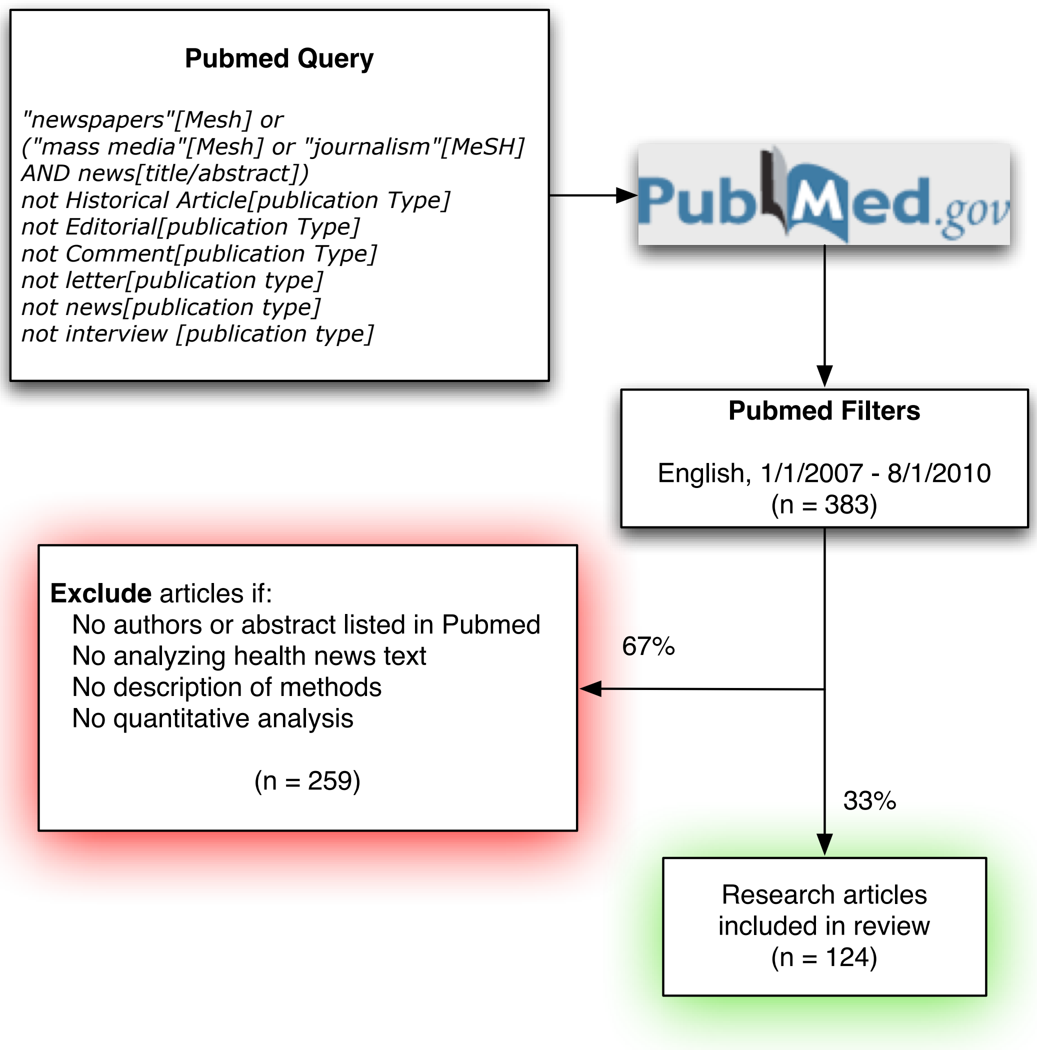


Figure 5. Literature review article retrieval

The PubMed query returned 383 research articles. Exclusion criteria (see page 27) removed 259 articles leaving 124[*]. Research articles were published in a variety of journals (*n*=77) including disease focused journals such as *Cancer Causes & Control*, health communications journals such as *Health Communications* and *Journal of Health Communication*, clinical medicine journals such as *Pediatrics,* population health journals such as *Women & Health*, and broad public health research journals such as the *American Journal of Public Health*.

Table 3 lists the journals with the most studies included in this review. In addition there were 14 journals that contributed 2 studies, and 52 journals that contributed 1 study. The large number of journals contributing only 1 or 2 studies suggests that while health news analysis studies are very common, there is no dedicated outlet for these studies (at least among the journals included in the Pubmed database).

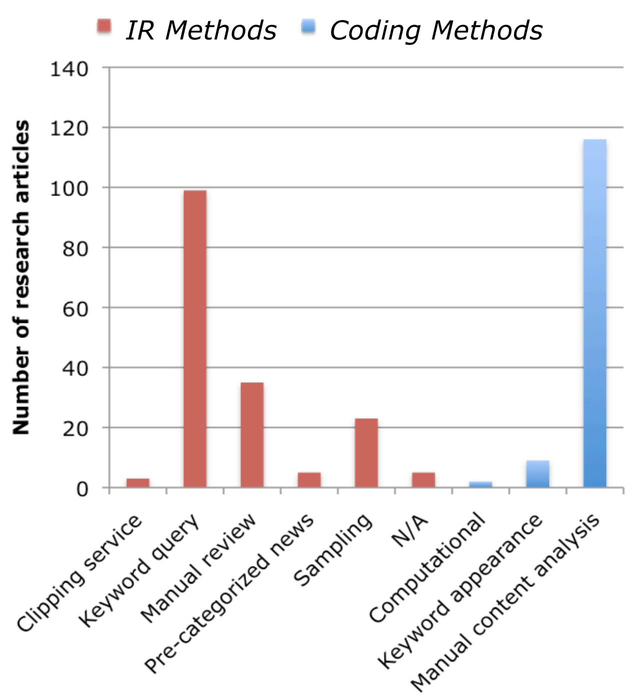
Journal Name	***n***
Journal of Health Communication	9
Social Science & Medicine (1982)	9
Health Communication	5
PLoS ONE	5
Patient Education and Counseling	3
American Journal of Preventive Medicine	3
Journal of the Royal Society of Medicine	3
Risk Analysis: An Official Publication of the Society for Risk Analysis	3
Tobacco Control	3

Table 2. Journals with the highest number of literature review research articles

[*] For the full list of articles refer to Appendix A on page 168

Research Methods

Of particular interest to this dissertation are the research methods most frequently used by investigators. A vast majority of the reviewed studies used simple keyword filtering to retrieve relevant health news coverage (*n*=99, 80%). Some retrieval methods, such as random sampling (*n*=23, 19%), were used in concert with keyword filters. Others methods, such as clipping services (*n*=3, 2%), and pre-categorized news articles (*n*=5, 4%) were used instead of keywords. Some researchers manually reviewed news coverage either as the sole method of retrieving relevant news or to ensure that news articles retrieved by other methods were in fact relevant (*n*=35, 28%). Most researchers used manual content analysis methods to code news articles (*n*=116, 94%), while few studies used any automated methods (*n*=9, 7%).



IR Methods	n
Keyword filter	99
Manual	35
Sampling	23
Pre-categorized	5
N/A	5
Clipping service	3

Coding Methods	n
Content analysis	116
Keywords	9
Computational	2

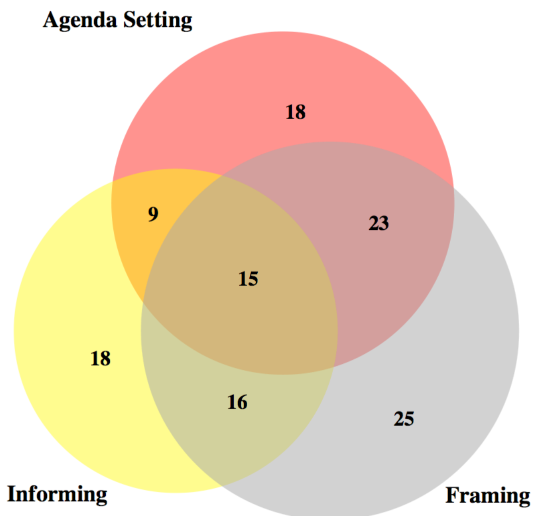
Figure 6. Literature review results: IR and coding methods

Investigators used a variety of analytical techniques in their research. However, many studies relied solely on descriptive statistics such as the direct comparison of frequencies of categorized news text (*n*=67, 54%). All studies included in this review to some extent reported descriptive frequencies, however the 67 studies referred to here relied solely on comparisons of frequencies. The remaining 67 studies used additional statistical analysis methods, many of which fall into the category of statistical hypothesis testing (see Table 3).

Statistical data analysis method	***n***
Chi-square test (Pearson's)	34
Regression analysis	12
Odds ratios	9
ANOVA (analysis of variance)	4
Fisher's exact test	3
Student's *t*-test	3
Time series analysis	2
Cochran's Q test	1
Cohen's kappa statistic	1
Gamma test	1
Mann-Whitney U test / Wilcoxon rank-sum test	1
McNemar test	1
Principle components analysis	1
Z-test	1

Table 3. Statistical analysis methods

Of the 124 research articles in the review, framing was studied the most (*n*=78, 64%), followed by agenda setting (*n*=65, 53%) and informing (*n*=58, 47%). Approximately 50% (*n*=63) of the studies investigated multiple media effects (see Figure 7).



Media Effect	*n*
Framing	78
Agenda Settings	65
Informing	58

(*Venn diagram generated by BioVenn*[*] [72])

Figure 7. Literature review results: distributions of media effects

The median number of news articles evaluated in the studies was 473. Two outliers evaluated a vastly higher number of news articles, Chau's study of the Australian media's coverage of physical activity[40] (84,136 news articles), and the agenda setting portion of Dong's study of news coverage of AIDS in Chinese[73] (607,440 news articles). Both analyses used keywords to code news articles.

[*] Available online at http://www.cmbi.ru.nl/cdd/biovenn/

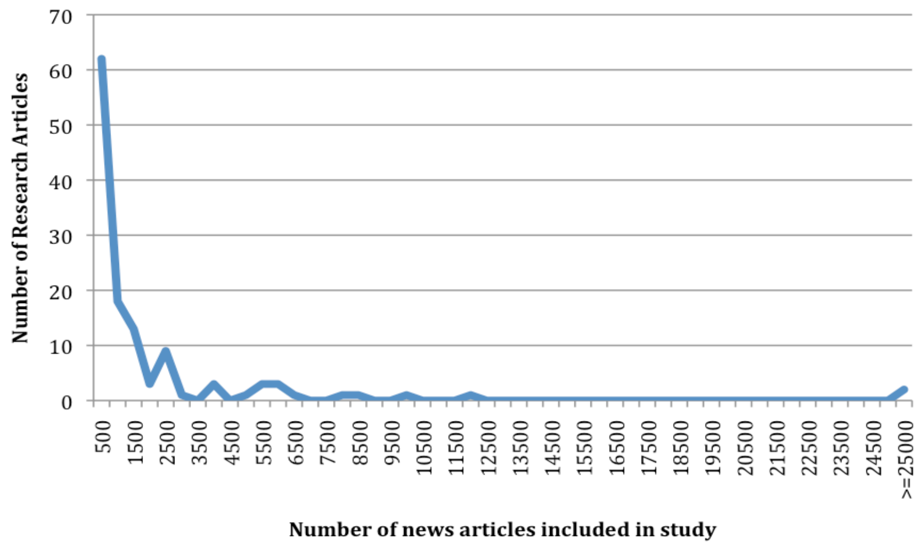


Figure 8. Literature review results: news articles evaluated in reviewed literature

Researcher interests

Researchers most frequently studied news coverage of disease and illness (*n*=36, 29%), followed by risks to health (*n*=22, 18%), and the media's coverage of accidental, violent, and self-inflicted injuries (*n*=12, 10%).

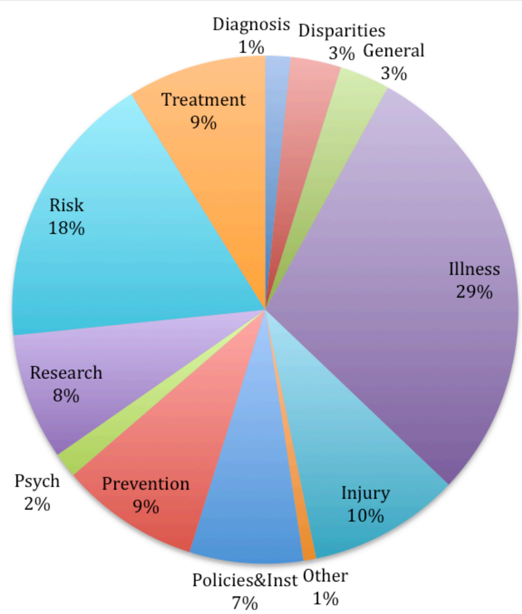
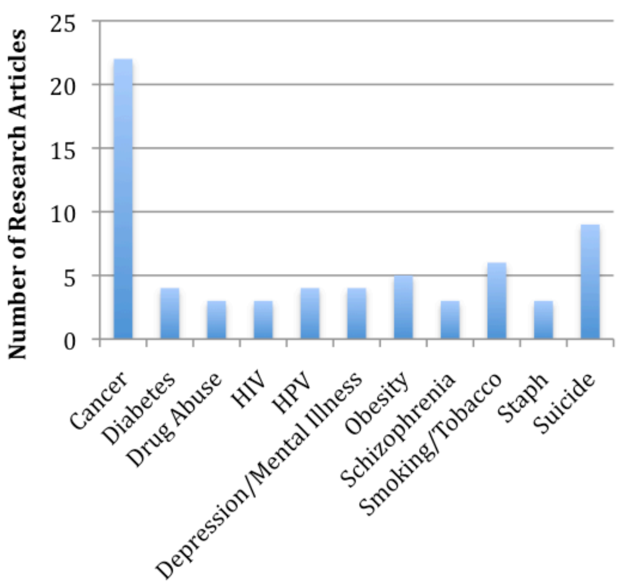


Figure 9. Health topic categories

Name	*n*	Health news topic studied
Disease / Illness	36	News coverage that discusses diseases and illnesses in broad terms
Risk	22	News discussing risks to good health
Injury	12	News of injuries (e.g. car crashes, violence, suicide)
Prevention	11	News discussing the prevention of illness and the maintenance of good health. (e.g. diet and exercise)
Treatment	11	News discussing treatments for disease
Research	10	Scientific research news
Policies & Inst	9	News of policies, campaigns and legislation by government and corporate institutions
Disparities	4	News of disparities in the health of populations
General	4	News coverage of broad health issues
Diagnosis	2	News of diagnostic tests
Psych	2	News of the psychological impact of poor health. (e.g. coping with a catastrophic diagnosis)
Other	1	

Table 4. Literature review: categories of health news investigated in studies

Cancer news was the most studied specific health issue (*n*=22), followed by news coverage of suicide (*n*=9), smoking/tobacco (*n*=6), and obesity (*n*=5).



Health Topic	*n*
Cancer	22
Suicide	9
Smoking/Tobacco	6
Obesity	5
Diabetes	4
HPV	4
Depression/Mental Illness	4
Drug Abuse	3
HIV	3
Schizophrenia	3
Staph	3

Figure 10. Top specific health news issues investigated by researchers

The majority of the studies evaluated non-episodic, long-lived news topics (*n*=88, 71%), but there were also a number of analyses that evaluated event driven, episodic news coverage (*n*=36, 29%). Examples of event driven health news evaluations included studying the media's coverage of the HPV vaccine[74-77] and MMR vaccination[78] controversies, hurricane Katrina news coverage[37] and the reporting in Europe of the H1N1 pandemic in 2009[79].

Most research articles evaluated news coverage from a single country (*n*=114, 92%). US news was the most frequent (*n*=52, 42%), followed by the UK (*n*=15, 12%), Australia (*n*=14, 11%), and Canada (*n*=8, 6%). Most studies evaluated news coverage that was published in countries where English is an official language (*n*=94, 76%). Only 9 studies (7%) evaluated news that spanned more than one country, and only 5 (4%) evaluated news on a broad international scale.

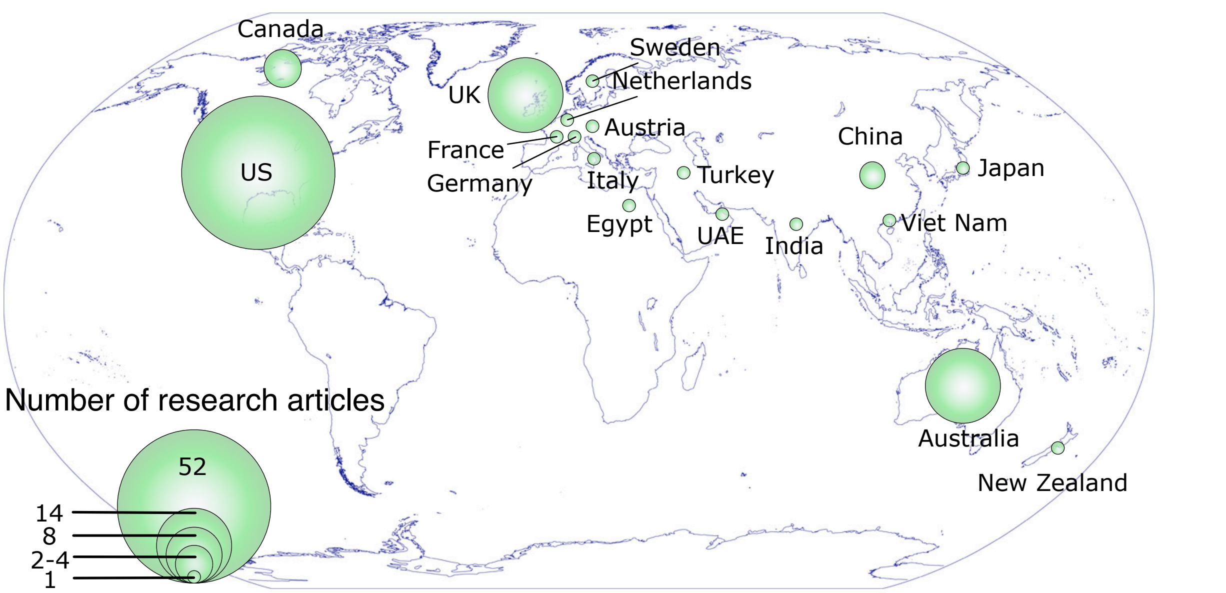


Figure 11. Frequency of research of news from different countries

Twenty-five research articles (20%) examined news coverage of health issues that had specific relevance to particular populations, or investigated other population features of health news coverage (e.g. target audience of news publisher, health disparities news coverage). Examples included Cohen's investigation of the differences in cancer coverage between general-audience news publishers and black news papers[80], cancer coverage mainstream and Korean online newspapers[81], and analysis of the dissemination of Canadian child maltreatment surveillance data[82]. Four of these studied evaluated news coverage that directly discussed health disparities[28-31]. Health news coverage discussing or relevant to the health of children and adolescents was the most frequently studied (*n*=15, 12%), followed by women (*n*=11, 9%).

Researchers often did not provide a clear assessment of the quality of the news coverage (*n*=37, 30%). Of the studies that did provide such an assessment (*n*=87) most gave a negative (*n*=48, 55%) or mixed (*n*=28, 31%) assessment of quality. Very few studies gave a positive assessment of health news coverage (*n*=11, 13%).

Health news population	**Research articles (*n*=25)[§]**
Children and adolescents	15
Women	11
Various ethnicities	4
Blacks	3
Men	2
MSM[**]	2
Korean Americans	1

Table 5. Research articles evaluating population sensitive health news

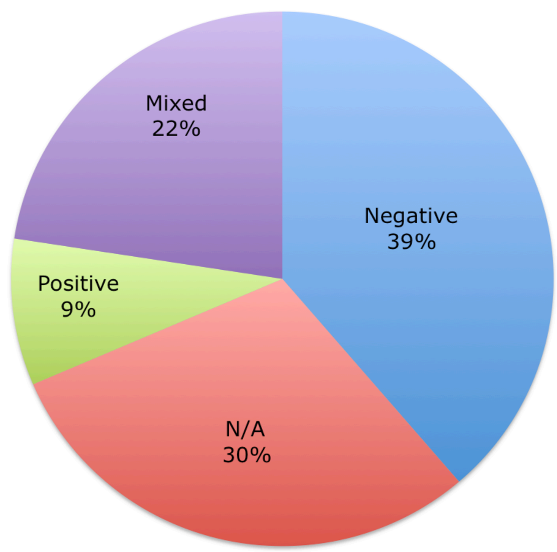


Figure 12. Quality of health news coverage

[§] 25 research articles in total. Some research articles investigated news discussing multiple groups.
[**] MSM = Men who have sex with other men

4.4. Discussion

Research methods – retrieval of relevant health news articles

Important to this dissertation are the methods that researchers use when investigating health news coverage. For the retrieval of relevant health news articles researchers relied heavily on keyword queries. This supports hypothesis 1A of this study.

Why do researchers rely so heavily on keywords to retrieve relevant health news articles? One possible answer is that the resources used most frequently to retrieve news coverage use keyword searching as the primary, or sole, querying interface. Lexis-Nexis, Factiva and Google news were the most common resources used to retrieve news coverage. These resources provide some filtering functionality for limiting search results based on metadata (e.g. publication date and publisher), but the primary querying mechanism is keywords and phrases.

To use more robust IR methods researchers would have to acquire corpora of health news coverage that meet their broadest inclusion criteria, and then apply IR methods to identify health news that is relevant to their research. Alternatively researchers could find and use a health news database that exposes more robust IR methods.

Acquiring a broad health news corpus and applying IR methods to identify relevant health news articles would give researchers the greatest control over the IR process, and would generate the most reproducible and reliable results. However, this approach requires expertise in computational and IR methods that the typical health news researcher may not have. Additionally, more extensive, dedicated computational resources would be required. For these reasons researchers will likely not find the process of acquiring a health news corpus and directly applying IR techniques to be a realistic alternative to keyword-based retrieval.

Alternative health news retrieval systems that provide more robust IR methods are difficult to find, or do not currently exist. However, later in this dissertation I will describe the design of a system that will support such features. This system, SalientHealthNews, and the methods and algorithms incorporated within the system, are based on the findings of this literature review, and the findings and algorithms from the classification study described later in this dissertation.

<u>Research methods – coding of health news articles</u>

Researchers almost exclusively used manual content analysis methods to code news articles. Very few researchers used automated content analysis methods, even for the simplest of coding tasks. This supports hypothesis 1B of this study. The almost exclusive use of manual content analysis is likely due to the fact that the majority of investigators studying the media's coverage of health approach this research from a social sciences perspective where manual content analysis is more prevalent.

<u>Research methods – research questions and analysis</u>

The choice of an appropriate quantitative analysis method depends on the research questions being asked and the data available for analysis. The studies in this literature review tended to have research questions that were combinations or variants of the following questions:

- Is health news coverage of category *X* (e.g. a health news framing, health topic, or news containing information of a particular type) more or less frequent than health news coverage of *Y*? [83-88]
- How does health news coverage of type *X* compare to health news of type *Y* on various measures (e.g. length, use of certain words, readability)? [88-90]
- How did health news coverage of category *X* vary over time? [91-94]
- How did the number of news articles of different categories vary based on publisher, geography, or other factors? [81,95,96]

The simplest quantitative analysis method that can be performed across all of these research questions is a direct comparison of frequencies of categorized news coverage. This technique, the simple comparison of descriptive frequencies, was the most frequently used analytical method in the reviewed studies. In approximately 50% of the reviewed studies this method was the sole analytical technique employed. This finding conflicts with hypothesis 1C, where I expected to find that the most frequently used analytical methods would be basic statistical significance tests such as Student's *t*-test and the chi-square test.

Statistical and analytical tests that are more robust than comparisons of frequency are often recommended because comparisons of frequency do not account for chance as many statistical tests do, and don't indicate the significance of differences that are found, or confidence in the frequencies that are reported. Why then did so many researchers only perform comparisons of frequencies? Some possible reasons are:

- The scope of the studies often did not justify the use of more robust statistical analysis methods
- The studies did not generate data that was appropriate for more robust statistical tests
- The sample sizes (number of news articles analyzed or that received various classifications) were insufficient for statistical testing

If a study of health news coverage is a preliminary investigation that is part of a larger study, and a broad descriptive characterization of health news coverage is all that is required, then descriptive frequencies may be sufficient. This is especially the case if the health news coverage that is available for analysis is not sufficient for more robust statistical analysis. This may be the case when the volume of health news coverage is very small or if other requirements or assumptions necessary for the use of more robust statistical methods cannot be met.

<u>Media framing</u>

Researchers more frequently investigated media framing, followed by agenda setting and informing. Framing studies being more frequent than agenda setting studies conflicts with part of the second hypothesis of this study, which posited that agenda setting studies would be the most frequently studied media effect. However, a closer look at the framing studies in this review reveals that these studies cover a very wide range of study types. Some framing studies investigated the general tone or slant of news articles, examples include Barnes' study of news coverage following Hurricane Katrina[37], and Scully's study of news coverage of skin cancer prevention[92]. Other framing studies examined aspects of health news coverage that help to frame a news story in a particular way. Two examples of this type of framing study are Kim's study of the framing of health disparities[31], and O'Hara's investigation into news media portrayals of eating disorders[97]. Still other studies categorized entire news articles based on broad framing categorizations such as health promotion, personal responsibility. Examples of this type of framing study include Andersson's study of media coverage of a diabetes prevention program[36], Henderson's study of the framing of news coverage of fast food advertisement regulations[25], and Hong's study of the framing of SARS in Chinese newspapers[98]. This wide range of framing study types supports the assertions of researchers such as Scheufele, Entman and D'Angelo regarding the loose and inconsistent definitions for news framing [61-63].

Informing studies were the least common. This supports part of the second hypothesis for this study, which posited that informing studies would be the least frequently investigated of the media effects. This was hypothesized because informing studies appear to be the most difficult and time-consuming type of media study type, requiring that news articles be read in depth searching for the presence of specific types of information in news text. Another reason may be that researchers favored exploring the appearance of various types of information in news articles for its potential to frame the news article, rather as a way of measuring the quality of information present in the news article.

Health topics and research questions

Researchers were interested in a wide range of health news topics and research questions. Research on media coverage of illness was the most frequent (*n*=36, 29%). Of the illness and disease news coverage studied, cancer was the most numerous (*n*=22). A close look at the cancer news analyses revealed that researchers were interested in a variety of cancer subtopics including various forms of cancer (e.g. lung, skin[92], cervical[76,77], prostate[41] and breast cancer[55]), cancer prevention[74,76], cancer treatment and outcomes[57], and cancer coverage about or directed towards specific populations[99,100].

One unexpected category that emerged during the review was analyses of the news coverage of *injuries*. Health news topics investigated in these studies were a combination of various health events that resulted in injury or death including suicide[60,94,101-105], drug use[106-108] and abuse of vulnerable populations such as the elderly[96]. Individually these topics appeared to be niche, and tied to very specific health events. For example, many of the suicide news coverage studies were a direct consequence of new guidelines set by the World Health Organization on how to report on suicide[4]. Future research should investigate if these health news topics persist and if the *injury* health news analysis category continues to interest researchers.

Episodic vs thematic news coverage

Researchers more frequently evaluated, by a margin of more than 2 to 1, non-event driven health news coverage over event driven health news. However, researchers have often found that the media tend to focus on unlikely but catastrophic events over more common but equally deadly risks[11]. This difference between what researchers most frequently study and what dominates news coverage may produce a gap resulting in health news coverage that is not evaluated. Future studies of health news analysis research literature should explore if such a gap truly exists. This is important because while it may be that the media should objectively cover certain types of health news more than others, it is also vital that researchers investigate what the media *is* covering, and how they are covering it.

How different populations are discussed in health news coverage

Twenty-five research articles examined news coverage that discussed particular populations or health disparities. Some of these research articles took special care to investigate how news coverage differs when specific populations were discussed, and how the media portray different populations when they are mentioned in conjunction with specific health issues.

Researchers found that to some extent media coverage discussing specific populations varies from other news coverage discussing the same health topic. In some cases this is to be expected, and can be a good thing. For example McDonnell, when looking at news coverage of Cancer in a leading Korean American newspaper versus a popular main stream newspaper, found that the main stream newspaper focused more on *"people, politics, and research"*, while the Korean American newspaper focused more on Cancer prevention. This made the Korean American newspapers a potentially good place for promoting community-focused interventions[81]. Stryker reported similar findings regarding a stronger focus on prevention when looking at a broader range of ethnic newspapers. Stryker also found that lower literacy levels were required to understand news coverage, and cancer awareness and education were discussed more in ethnic newspapers compared with main stream newspapers[88]. Cohen, while looking at Cancer news coverage in black newspapers compared to general-audience newspapers corroborated the findings of both Stryker and

McDonnell. Cohen also found that Cancer coverage in black newspapers had more local relevance, and personal mobilizing information than the general audience newspapers. These results can be regarded as positive in that they show that media coverage in ethnic newspapers provides news coverage of cancer that is especially relevant for the target news consumer, and promotes prevention.

Researchers also found that news coverage tends to reinforce existing social stigma or stereotypes concerning the populations under consideration. Bengs, in his study of how portrayals of depression in Swedish newspapers vary by gender, found that news stories about depression often presented *"relatively gender-stereotyped portrayals of women and men"* [109]. Gough had similar findings in the context of dieting, where he found that *"diet* [in the analyzed news coverage] *continues to be construed as women-centered (hence 'unmasculine')".* Gough goes on to say that *"men's relation to food and health is framed belies the continued dominance, in the media at least, of hegemonic masculinities*"[110]. Halpin also found that the media continue to emphasize stereotypical masculinity in news coverage of prostate cancer. Halpin stated that the *"meta-narrative was one of traditional masculinity, in which courage, money, success and virility trumped self-health and/or discourses about how it feels and what it means to have* [prostate cancer]*"* [41]. Finally Schwartz found that news coverage of methamphetamine use by heterosexual and gay men stigmatized both populations in different ways, gay men were more likely to be discussed in terms of sexual health issues, while crime was the more dominant context in which heterosexual men were discussed[107].

These findings illustrate the varying ways that the news media discusses well covered health issues such as cancer, mental illness and diet differently when particular populations are also being discussed. These results suggest that further study in this area is warranted.

<u>News sources</u>

Most of the studies in this review examined news coverage that was published in the US (*n*=52) or in other English speaking countries. This was not surprising due to the exclusion of research not published in English. Also contributing to this may be that journals containing analyses of news published in other countries may not be included in PubMed. Future literature reviews in this area may need to focus on research that investigates news coverage from non-English news publications.

<u>News Quality</u>

Only 9% of the reviewed studies reported that health news coverage was of good quality, while 39% reported that the quality of health news was inadequate. These findings indicate that shortcomings in the media's coverage of health persist. This supports hypothesis 3 of this study, and is further evidence that continued monitoring, evaluation, and improvement of health news coverage is needed.

Researchers provided numerous reasons for the poor quality of health news coverage. One general concept that concisely describes many of these problems is bias. Entman described 3 specific types of bias that often occur in news coverage, they are the bias that appears due to a distortion or falsification of facts (*distortion bias*), bias that favors one side of an argument over another (*content bias*), and bias that due to the mindset and preconceptions of the journalist creating the news content (*decision-making bias*)[111].

Distortion bias can occur when journalists discuss health in a dramatic and sensational way, or when the risk of health problems that are fantastic or rare are emphasized relative to the risk of more common but less sensational health problems. Both of these manifestations of bias can lead to inaccuracy in the health information contained in health news coverage[112]. One example of this form of bias can be seen in Akamatsu's study of news coverage of diabetes mellitus in Japanese newspapers[113]. Akamatsu found that when selecting research studies to discuss in news coverage, journalists preferred research that described scientific and medical breakthroughs. Another example of distortion bias is presented by Berry in his study on the role that risk plays in the construction of health news and the selection of health topics for reporting[114]. Berry found, over a span of 4 years, that the prominence of sudden acute respiratory syndrome (SARS), and mad cow disease were greater than, or comparable to more common but less sensational diseases and health risks such as cardiovascular disease, HIV/AIDS, and diabetes[114].

Content bias can be seen in cases where journalists attempt to personalize news coverage by discussing the stories of individuals. By emphasizing individual-level explanations of health problems journalists risk over prioritizing individual responsibility at the expense of societal responsibility. This type of bias can be seen in Kim's study of the news coverage of racial and ethnic health disparities in US newspapers[31].

Another form of content bias occurs when the sources quoted and referenced in health news coverage do not fairly represent the various actors, stakeholders and viewpoints surrounding a health issue. Brown provides a good example of this problem in his study on the reporting of the use of medical records in research[115]. Brown found that while numerous sources claimed to speak on behalf of the public's interest regarding privacy concerns, there was no direct evidence that these sources truly communicated the public's opinion, or that the public's views were actually represented in the news coverage.

Decision-making bias occurs when journalists reflect the biases, stereotypes, prejudices and misconceptions present in the population they serve. Numerous examples of this were reported in the reviewed literature. Examples included the news media's stigmatization of mental illness[116-118], and sexuality[107], and the media's propagation of generally held misconceptions regarding the inherent behavior of men and women in relation to various health issues[41,109,110].

Even when none of the biases described by Entman occur, problems can still arise simply due to the inherent challenge of communicating scientific and medical information clearly, accurately and fairly. Health news is particularly difficult to report because journalists have to juggle the need to inform the public about health issues with the practical needs of news publishers to attract an audience and be profitable[67,119,120]. This problem is only worsening as news organizations cut back on journalists because of declining profits[59].

Limitations

One limitation of this study is that only research published in English was included. There may be relevant research published by non-English researchers or published in non-English journals. Another limitation is that this review only covers the period from January 1, 2007 to August 1, 2010. This period of time gives a good snapshot of recent studies, but does not cover enough time to identify changes in researcher interests and the use of methods over time. Another limitation is the exclusive use of Pubmed for retrieval of research articles. This may result in the exclusion of relevant research articles published in journals not indexed by Pubmed.

4.5. Summary

In this chapter I described a literature review of research articles that evaluated health news coverage. The aim of the review was to identify the methods and research interests most frequently studied by health news investigators. Most relevant to this dissertation were the findings that researchers rely primarily on keyword filtering and manual content analysis methods for retrieving relevant health news coverage and for scoring and coding of news coverage. Other findings included that (a) researchers most frequently found that health news coverage was inadequate, (b) researchers more frequently used simple descriptive frequencies compared to other statistical methods, and (c) news framing was the most frequently studied media effect. These findings provide a clear picture of what health news researchers currently investigate and how they perform their investigations, and were used to inform the classification study described in Chapters 7 and 8, and design the SalientHealthNews system described in Chapter 9. In the next chapter background information on computational methods relevant to the research described later in this dissertation will be discussed.

Chapter 5. BACKGROUND: COMPUTATIONAL METHODS

In this chapter I discuss computational concepts and algorithms that were used in the projects discussed later in this dissertation. In Section 5.1 I describe basic concepts from the areas of information retrieval and document classification. In Section 5.2 specific classification features including *n*-gram language modeling, topic modeling and readability are described. In Section 5.3 I discuss ways of combining classification features to produce a binary classifier. In Section 5.4 I discuss methods for evaluating classifiers. Section 5.5 discusses two classification systems, BoosTexter and an SVM/LDA classifier, that have been shown to perform well on text classification tasks, and were compared to the classification system developed for this dissertation.

Before diving in to what is a detailed discussion of numerous complex computational methods and concepts, readers not familiar with basic information retrieval and document classification concepts may find the following glossary of terms helpful.

Glossary of information retrieval and document classification terms

Cases/items – In the context of information retrieval, and classifier development, the terms *case* and *item* refer the things being classified or retrieved.

Corpus - A collection of documents.

Document - In information retrieval and text classification, documents are the unit of text being classified or retrieved. For this dissertation a document is a news article.

Feature **–** A classification feature is an attribute associated to the items being classified that can be used by a classification algorithm to predict the class/category of an item, or can be used to score the item in some way.

Gold standard - A set of correctly classified items typically used to train and evaluate the performance of a classifier or information retrieval method.

Index, indexing - Information retrieval methods often create a model of the corpus being searched. This model is then subsequently used when information retrieval tasks are performed. This model is referred to as the index. The act of creating the model is referred to as indexing the corpus.

Positive/negative (cases or items) - Positive or negative refer to the binary classification received by a case/item.

Query - A set of search parameters that can be submitted to an information retrieval method. What specifically constitutes a query is based on the information retrieval method.

Significance (statistical) – Statistical significant refers to a result that is not likely to have occurred by chance.

Test/validation set - Similar to a gold standard, this is a set of correctly classified items. A test/validation set is used solely to estimate the performance of a classification or information retrieval system.

Training set - A set of cases/items used to train a classification system. A training set is typically comprised of positive and negative cases. The precise size of a training set, and the distribution of positive and negative cases often varies based on what is available and what is required by the classification system.

Type I/II errors – A type I error is synonymous with a false positive. This is when an item is incorrectly classified as a positive case. A type II error is synonymous with a false negative. This is when an item is incorrectly classified as a negative case.

5.1. Classifier design

Kitzinger and Gamson's work on media templates and media packages (discussed in Section 3.2) suggest that patterns exist in the text of news coverage[64,66]. A simple example of this is the repeated appearance of specific terms and phrases in news articles that discuss particular topics. For example it is reasonable to assume that the term *obesity* and the phrase *body mass index* will appear more frequently in news coverage about obesity compared to news discussing other topics. The literature review discussed in the previous chapter shows that health news researchers rely heavily on this assumption when they attempt to retrieve news coverage relevant to a particular health topic. But as discussed earlier, relying on keywords for retrieving relevant documents can be problematic. Problems that arise when relying on keywords include:

- There may be news articles relevant to the topic of obesity that do not include the term *obesity*
- There may be news articles that contain the term *obesity* but only discuss the topic of obesity in passing
- Simple keyword searching lacks a way to fine tune or rank results based on relevance
- Manually defining keyword queries that can, with high accuracy, retrieve documents that are relevant to complex concepts such as environmental risk or biological causality can be extremely difficult

Numerous information retrieval and classification methods have been developed that can address these problems. The design of such classification methods continues to be a very active area of research in computer science. Figure 13 illustrates the various stages involved in designing a classification system.

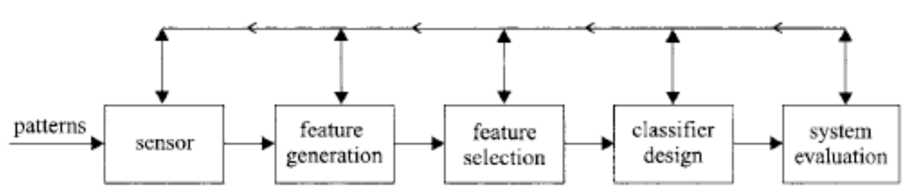


Figure 13: The design stages of a classification system (Theodoridis 2006[121]*)*

In the first stage, the *sensor* stage, the mechanism for measuring and collecting raw data is defined. In the context of studying health news this stage is the basic collection and storage of the universe of news content that is considered for analysis (e.g. all news published by a set of news sources over some period of time). The *feature generation* stage describes the algorithms and calculations used to generate the data upon which classifications will be based. For news text simple features include news article length, and the sequence of terms in the news article text. The *feature selection* stage is where features are evaluated to discover those that are the most appropriate for the classification task. This phase is most useful when it is not known what features will produce the best classification performance, and so a number of different features are explored. *Classifier design* is where a strategy is defined for combining classifier features to produce a final classification. Lastly, a means of evaluating the performance of the classification system is defined in the *system evaluation* stage.

Classification design is typically a very iterative process, as illustrated by the feedback and bi-directionality of the connections in Figure 13. This is very much like other iterative and incremental design models like the spiral model[122], and agile software development[123], that promote constant reevaluation of design choices based on the knowledge learned throughout the design, implementation, and evaluation of a system. This is in contrast to the much maligned waterfall design method that is too strict to be practical or successful for complex design projects[124].

5.2. Classification Features

Numerous classification features were considered for this dissertation. Those with the best potential to contribute to the classification of health news coverage based on health topic and news framing were selected for implementation and evaluation. In this section I will describe the classification features that were selected.

Reliance on word frequency

Text classification features often rely directly or indirectly on word frequency. The direct reliance on word frequency often takes the form of automatically identifying statistically salient terms and phrases in a corpus or document relative to another corpus or document, or in the counting of word types to identify groups of words that are more frequent and therefore more representative of a text relative to other texts. Direct use of word frequency can produce very good results, especially when specific words and phrases can be relied on as an accurate guide to what a text is about. In addition, direct use of word frequency is one of the most logical and easy to understand approaches to text classification and information retrieval. However there are drawbacks. Relevant but extremely infrequent terms are often not identified by these methods, and the direct use of word frequency relies on implicit assumptions regarding the distribution of words in a text and corpus, assumptions that if incorrect may result in poor performance.

Numerous modern text classification and information retrieval techniques rely indirectly on word frequency to derive higher quality information over which text can be analyzed, retrieved and classified (see the following sections on language modeling and topic modeling for specific techniques). These methods often outperform more direct methods (e.g. keyword searching and identifying statistically novel terms), however these techniques have their own drawbacks. They often require large amounts of training data to prepare them for classification and information retrieval tasks, they tend to be more computationally expensive than more direct methods, and they are less intuitive for individuals that do not have extensive experience in the area of text data mining.

To mitigate some of the problems that arise when relying solely on either direct or indirect use of word frequencies, the classifiers developed in this dissertation use a mixture of methods, combining these techniques in a way that potentially maximizes information retrieval and classification performance.

Novel term detection

The literature review in the last chapter revealed that health news investigators most frequently use of keywords for identifying relevant health news coverage. However, as already stated, keywords are problematic. One shortcoming of querying with keywords is that keyword queries need to be manually defined. When searching for health news relevant to a complex concept such as framing, identifying a keyword query that will produce accurate results can be very difficult. A better approach is to automatically identify keywords and phrases using statistical methods.

The online retailer Amazon.com uses a method named Statistically Improbably Phrases (SIPS) to provide customers with a set of phrases that distinguish a book from other books in the Amazon catalog[125]. A book's SIPS are automatically discovered by calculating the *z*-score (see Equation 1) of the frequency of terms and phrases appearing in a book compared to the average frequency for that term or phrase across all books in the Amazon corpus. Since *z*-score measures the difference between an observed datum and a population mean, the terms and phrases with the highest *z*-scores for a book tend to be the terms and phrases that most differentiate that book from the rest of the corpus. Once discovered, SIPS can be used like keywords to query for other books that are similar to the text that produced the SIPS.

A similar method, CKA, uses the chi-square test to identify terms that appear significantly more or less often in a text, versus the text's corpus. One example of CKA in use is Seale's investigation of differences in the language used by men and women in online cancer support groups[126].

The CKA and SIPS methods are analogous to the information retrieval method tf-idf (term frequency-inverse document frequency), which is used to give appropriate weights to all possible keyword-document combinations in a corpus[127]. The tf-idf weight of a term increases proportional to the frequency of that term (*tf*) in a given document offset by the frequency of that term throughout the corpus (*idf*). The result is that the "best" search terms are those that appear very frequently in relevant documents, but appear less frequently throughout the rest of the corpus (see Equation 2). In this dissertation automated keyword and short phrase features were identified using a combination of SIPS and tf-idf techniques.

$z = \frac{x - \mu}{\sigma}$	*Equation 1. Z-Score* $\boldsymbol{x}$ = observed score $\boldsymbol{\mu}$ = population mean $\boldsymbol{\sigma}$ = population standard deviation
$tf_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}}$ $idf_i = \log \frac{N}{n}$ $(tf \cdot idf)_{i,j} = tf_{i,j} \times idf_i$	*Equation 2. Term frequency - inverse document frequency* ***$tf_{i,j}$*** *is the term frequency for term* ***t_i*** *in document* ***d_j****.* ***$n_{i,j}$*** *is the number of times* ***t_i*** *appears in* ***d_j*** *.* ***N*** *is the total number of documents in the corpus,* ***n*** *is the number of documents* ***t_i*** *is in.*

N-gram Language Modeling and Perplexity

N-gram language modeling is a method for capturing regularities in word use throughout a corpus. In an *n*-gram language model, probabilities are calculated for the appearance of words and sequences of words in a corpus. Typically bi-grams or tri-grams (i.e. sequences of length of 2 or 3) are used[128]. Probabilities are often smoothed to better estimate probabilities when there is insufficient data[128]. The resulting probability distribution over all *n*-grams in a document or collection of documents is the *n*-gram language model. The model of one document or corpus can be compared to the model of another to estimate the similarity between the model sources. Any method that can measure the difference between probability distributions can be used to measure the difference or similarity between two *n*-gram language models. Examples include perplexity, Kullback-Leiber divergence and cross-entropy[129,130]. Perplexity was selected for use in this dissertation because of its successful use in a prior obesity news classification study (described in detail in Chapter 6).

$PP(s_1,s_2,\ldots,s_k) = 2^{H(s_1,s_2,\ldots,s_k)}$	Perplexity of a text in terms of entropy (*H*), where the text is defined as a sequence of sentences ($s_{1\text{-}k}$).
$H(s_1,s_2,\ldots,s_k) = -\sum_{i=1}^{k} P(s_i)\log_2 P(s_i)$	Entropy of a text as a function of the probabilities of generating the sentences in that text based on the probability distribution (*P*) of a training text.

Equation 3. N-gram language model perplexity

When applied to *n*-gram language models, perplexity can be interpreted as a measure of the uncertainty of predicting sequences of words in a corpus given the knowledge of the probability of sequences of words in a training corpus. Low perplexity indicates that the text used to generate an *n*-gram language model has an increased ability to predict word sequences in test documents while higher perplexities indicate a decreased ability to predict word sequences.

N-gram language modeling can be used to identify documents relevant to a topic by first creating a language model for a set of positive training cases, then calculating the perplexity of that model relative to language models of test cases. The documents of language models with lower perplexities can be regarded as more similar to the positive training cases than documents with language models that produce higher perplexities.

A preliminary study evaluating the use of *n*-gram language modeling to differentiate general news, health text and obesity news is reported in Chapter 6. That study strongly suggested that *n*-gram language modeling would be a good classification feature for health news analysis[131].

Topic Modeling

Information retrieval methods such as keyword searching, and *n*-gram language modeling, treat documents as nothing more than sequences of terms. Topic modeling is an alternative approach that attempts to map documents a set of abstract topics, where the number of topics is significantly smaller than the size of the corpus vocabulary. When successful, topic modeling algorithms generate condensed topic spaces where the relationships between documents and terms can be searched, and where the "noise" associated with the clutter of terms irrelevant to the broad concepts that span the corpus are diminished.

An early effort in the area of topic modeling was latent semantic analysis[132]. The latent semantic analysis (LSA) topic space is based on the frequency of occurrence of terms in a corpus' documents. Term frequencies are stored in a simple occurrence matrix X, where one dimension of the matrix is documents and the other is terms. Indexing a corpus entails the calculation of the singular value decomposition (SVD) of the occurrence matrix X. The result of the SVD is two orthogonal matrices, T_o and D_o, and a diagonal matrix S_o. When multiplied, T_o , S_o , D_o produce the original occurrence matrix X (see Figure 14 a). In LSA instead of using an exact SVD (one where the product of T_o , S_o , D_o reproduces the original occurrence X), a lower rank approximation, T , S , D, is calculated (see Figure 14 b). The variable K is used to define this lower rank. The theory is that an approximation will merge together related terms, producing K topics over which all documents and terms in the corpus can be mapped. These K topics are potentially a denser conceptual space that contains less noise than the exact SVD and therefore performs better for document retrieval.

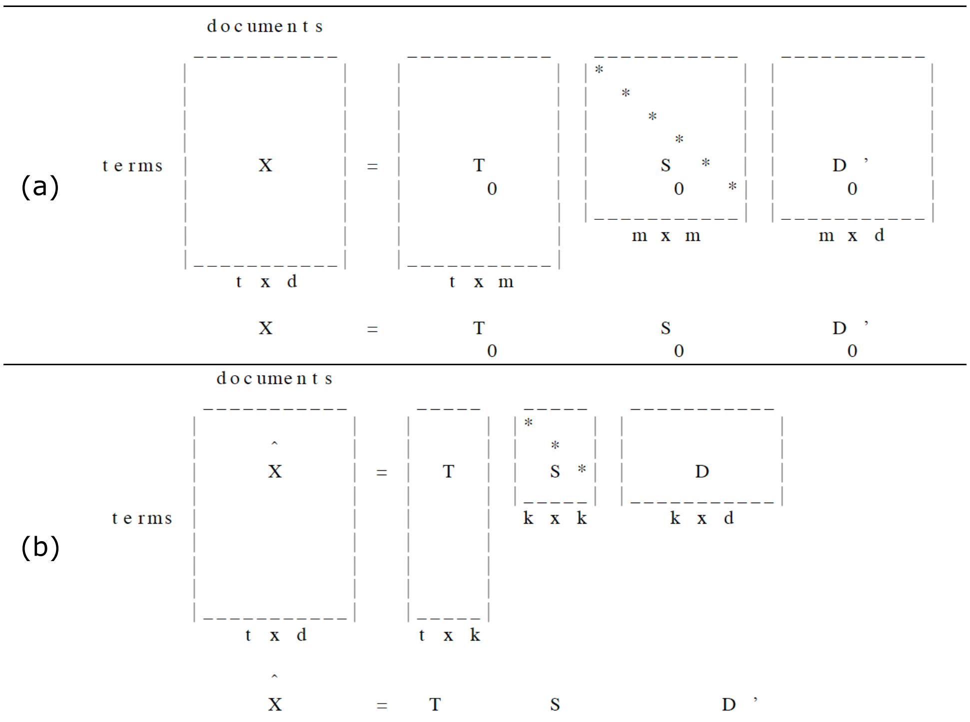


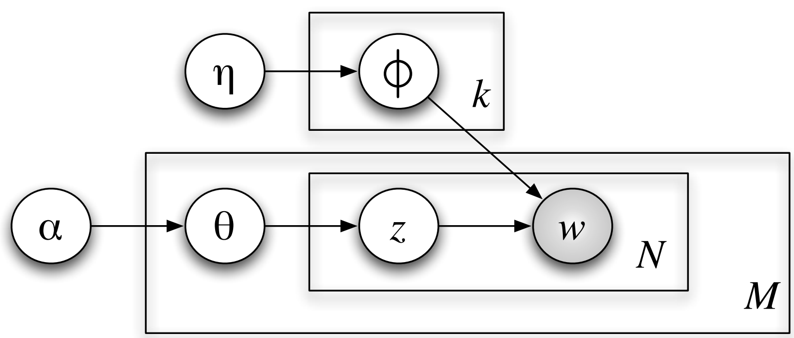
Figure 14. LSA (a) exact SVD, and (b) approximate SVD (Deerwester 1990[132])

The way to search for documents in LSA is to treat a query (a list of search terms) as a pseudo-document composed solely of the query terms. A pseudo-document vector is then constructed by calculating the centroid of the query term vectors in *T*. Similarity between the query and a test document from the corpus is estimated be calculated the dot product of the query's pseudo-document vector and the vector of a document in the corpus; the smaller the dot product, the greater the similarity between the query and the test document.

Topic modeling has seen many improvements and alternatives since LSA was proposed. Examples include Probabilistic Latent Semantic Indexing[133], mixture of uni-grams model, Random Indexing[135] and Latent Dirichlet Allocation[136]. Perhaps the most widely used technique, and the one used in this dissertation, is Latent Dirichlet Allocation (LDA).

In LDA, documents are modeled as random mixtures of latent topics, where topics are probability distributions over words. This later quality of the model leads to its being characterized as a generative model (a model that is capable of randomly generating observable data). Other examples of generative models include hidden Markov models and naïve Bayes. Generative models are advantageous because unlike non-generative methods (e.g. LSA), LDA can produce examples of sets of words that are representative of the topics in the model.

Figure 15 is a graphical illustration of how LDA models a collection of documents. The boxes in this illustration are "plates" and represent replicates that are repeated the number of times indicated at the bottom right of the box. In LDA, indexing a corpus entails estimating the probability distributions in ϕ and θ. This is a problem of Bayesian inference. Blei's original LDA implementation used a variational Bayes approach to estimate these distributions, but other approaches such as Gibbs sampling[137] are often used instead.



Key: N_i = The number of words in the ith document
M = The number of documents in the corpus
α, η = LDA hyper-parameters
k = The number of topics in the model
$\phi = k \times V$ matrix (V is the size of the vocabulary) where each line is the probability distribution for words given a specific topic, sometimes represented using β
$\boldsymbol{\theta}_i$ = The topic probability distribution for the ith document
$z_{i,j}$ = The topic of the jth word of the ith document
$w_{i,j}$ = A specific word in a document

Figure 15. Graph representation of LDA using "plate" notation (Blei 2003[136])

Once the probability distributions are available there are numerous ways to perform classification and information retrieval tasks. For example the similarity between different words or between documents can be estimated by comparing their probability distributions over the model topics, represented in Figure 15 as $\boldsymbol{\theta}_i$. As with the *n*-gram language modeling method described earlier, this can be done using any measure that compares probability distributions. Examples include Kullback-Leiber divergence[130], cross entropy and perplexity[131].

One method for calculating the similarity between a query and a document, possible only because of the generative quality of LDA, is to calculate a query's likelihood[138]. Scoring documents based on query likelihood entails the calculation of the probability of a document model D generating the query Q (Equation 4a). This calculation requires the posterior estimates $\hat{\theta}$ and $\hat{\phi}$ for $\boldsymbol{\theta}$ and ϕ (Equation 4b).

$P(Q \mid D) = \prod_{q \in Q} P(q \mid D)$	(a) basic query likelihood
$P(q \mid D) \equiv P_{LDA}(w \mid d, \hat{\theta}, \hat{\phi}) = \sum_{z=1}^{K} P(w \mid z, \hat{\phi}) P(z \mid \hat{\theta}, d)$	(b) LDA probability of a document generating a word (refer to the key in *Figure 15*)

Equation 4. Probability of a document generating a query in LDA

Another advantage of using LDA is that it easily generalizes to new documents without the need to re-index the entire corpus. This means that documents not in the originally indexed corpus, for example pseudo-documents that are created for document retrieval tasks, or additional documents acquired after indexing, can be easily added to the index for searching or querying.

Readability

The goal of readability research is to match reader to text[139]. This does not refer specifically to the content of text, but to the ease with which text can be read and comprehended by the reading audience. Readability has been shown to vary in health news coverage based on health topic[74] and on target reading population[88]. This makes readability a potentially effective classification feature for health news.

There are numerous attributes of a document that can impact its readability including the style of the writing (technical vs. narrative, active vs. passive voice, first vs. third person), vocabulary, linguistic structure, and even the formatting of the text and the use of images[140]. This can make the measurement of text readability difficult and time consuming. However, numerous formulae have been developed that provide easy estimates of the readability of text. These formulae typically rely on straightforward text features such as syllable, word and sentence counts. The formulae are often calibrated to reading levels of children in the United States. Table 6 lists some of the more commonly used readability formulae.

Despite the common and accepted use of these formulae, they have many shortcomings. Most readability formulae (all listed here) rely heavily or exclusively on word, sentence, and document length, with little or no regard to the domain of the content being measured. For example text from some domains, such as medicine and biology, exhibit heavy use of abbreviations and other word and phrase shortenings[141-143] that may tend to incorrectly produce readability scores that indicate that these texts are easy to read and comprehend. Readability formulae also neglect other factors that may influence the comprehensibility and learnability of text, including grammatical complexity and word usage[140], the background and familiarity the audience has with a content’s domain[144], and the density of concepts in a text[139]. In fact, these formulas can’t even differentiate between scrambled and coherent text.

Readability Formula	Score	Formula
FLES - Flesch Reading Ease[145]	Approximately between 0 and 120; 0 is best understood by US university graduates, 120 is easily understood by an 11-year-old student.	206.835 – 84.6 *wl* - 1.015 *sl* *wl* = avg word length (syllables) *sl* = avg sentence length (words)
Flesch-Kincaid Grade Level[146]	US grade level	0.39 *sl* + 11.8 *wl* – 15.59 *wl* = avg word length (syllables) *sl* = avg sentence length (words)
ARI - Automated Readability Index[146]	US grade level	4.71 *wl* + 0.5 *sl* – 21.43 *wl* = avg word length (chars) *sl* = avg sentence length (words)
SMOG - Simple Measure of Gobbledygook[147]	US grade level	$1.043\sqrt{30 \times \frac{p}{s}} + 3.1291$ *p* = # polysyllabic words *s* = # sentences

Table 6. Commonly used readability formulae

In spite of these shortcomings readability formulae have been consistently validated in numerous studies[144,146,147]. This, along with their ease of calculation, has led to their continued use in measuring the readability of all types of content including news text, health communications materials and educational materials[89,148,149].

Word type and sentiment

Sentiment analysis is the process of scoring text, based on the attitude, opinion, and emotion conveyed in the words and phrasing of the text. Sentiment scores may simply represent the polarity of the text (e.g. positive or negative opinion), or may estimate more subtle emotional traits such as objectivity, imagery and sexuality. Because many health news analyses code for the sentiment or tone of news coverage[32,37,53,74,150-152], sentiment features were used as a classification feature in this dissertation.

Sentiment analysis systems are typically developed using one of two techniques, a “bag of words” dictionary method, or machine learning approaches. Machine learning approaches use algorithms such as naïve Bayes and support vector machines to automatically classify the sentiment of text. These methods have been shown to perform better than dictionary approaches[153]. Unfortunately machine-learning algorithms typically require large amounts of training data, making them difficult to develop.

Dictionary methods require the construction and use of dictionaries that map words to scores of their emotional qualities. Sentiment dictionaries can be constructed manually using human coders[154], or automatically based on the linguistic characteristics of words[155]. Using a sentiment dictionary, a text can be scored by summing or averaging the sentiment scores of the words that appear in the text.

Two dictionaries that score words on dimensions of sentiment are the Whissell’s Dictionary of Affect in Language[154], and the Linguistic Frequency and Word Count dictionary[156]. The Dictionary of Affect in Language (DAL) is a word list of 8742 common English words rated from 0 to 3 on the dimensions of *pleasantness*, *activation* and *imagery*. The strength of the DAL is in its focus on these 3 affect dimensions. The Linguistic Frequency and Word Count (LIWC) dictionary maps over 4500 English words to one or more word categories. Some of the LIWC categories indicate when a word has sentimental connotations. These categories mostly fall under the LIWC category of *Affective Processes*. The strength of the LIWC dictionary lies in the breadth of its categories. Examples include various social process categories such as *family*, and *friends*, cognitive categories such as *tentative*, and *certainty*, and personal concerns categories such as *achievement*, and *leisure*. The breadth of these categories allows for the development of more complex scoring and categorization of documents. For examples of scores and categorizations from both the LIWC and DAL dictionaries refer to Appendix C on page 195.

5.3. Combining Features

Classification features must be aggregated to produce a classification system. Numerous computational methods can perform this task. Most can be characterized as supervised, semi-supervised, or unsupervised machine learning algorithms.

Supervised, unsupervised and semi-supervised classification

Supervised machine learning algorithms learn how to classify items by examining the traits of items in a training set. Examples of supervised machine learning algorithms include support vector machines[121] and naïve Bayes classifiers[157]. Unsupervised machine learning algorithms automatically detect similarities between unlabeled items, and use those similarities to group together similar items. Perhaps the most frequently used unsupervised method is cluster analysis. Semi-supervised methods use both labeled and unlabeled data to learn how to categorize new items.

Supervised machine learning methods typically outperform other methods. However, such methods often require non-trivial amounts of training data to achieve optimal performance. When sufficient training data is not available, unsupervised, or semi-supervised methods are used.

Semi-supervised methods use a limited set of labeled training cases to learn or constrain classification rules. These rules are then applied to a much larger set of unlabeled items, which serve to further refine those rules in a manner similar to unsupervised classification methods. The use of a limited set of training cases and a much larger set of unlabeled cases makes semi-supervised methods more appropriate for interactive querying and classification tasks, or for situations where a small amount of training data can be acquired and used to boost the performance of an unsupervised method.

Two examples of semi-supervised methods are constrained clustering[158], and co-training[159]. Constrained clustering is a modified version of typical unsupervised clustering where constraints are placed on how labeled items are grouped. Specifically examples of constraint rules are *must-link* and *cannot-link* constraints, *Must-link* constraints indicate that certain labeled items must ultimately be clustered into the same group. *Cannot-link* constraints define labeled items that must not end up in the same group. These constraints act as a guide for the groups ultimately generated by the clustering algorithm.

In co-training multiple classifiers are trained using ideally independent sets of features, where each set of features is sufficient for correct classification. The classifiers are initially trained on a small set of labeled items. The strongest classifications of unlabeled items are used to iteratively train and classify additional unlabeled items.

K-means clustering

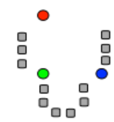
K-means clustering was selected for use in this dissertation. *K*-means is a basic and widely used unsupervised classification method[160]. Figure 16 illustrates the basic *K*-means algorithm. In the algorithm each item being clustered is represented as a vector of numbers corresponding to some set of clustering features. The algorithm follows a very basic expectation maximization (EM) cycle, making it a member of the family of EM algorithms that exist in statistics and computer science.

EM algorithms iteratively repeat two basic steps, the expectation step and the maximization step, in order to identify the parameters of some statistical model. In the expectation step the algorithm applies an existing estimate of the parameters it is trying to identify and uses them as if they were the final, correct estimates for the dataset. In the maximization step the algorithm calculates a new set of parameters that maximizes the parameters estimated in the expectation step.

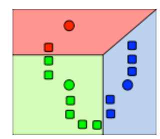
In *K*-means, after an initial step where the geometric centers of the *K* clusters being calculated are randomly selected, the algorithm proceeds by assigning all items to the nearest defined cluster center (this is the expectation step), then recalculating the cluster centers based on the items in that cluster (the maximization step). These steps are repeated for some number of pre-specified iterations, or until no significant change in cluster assignments or cluster centers are observed.

K-means clustering algorithm
(images taken from [161]*)*

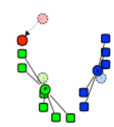
1. Algorithm initialized by randomly selecting *K* items to be the "means" (the red, blue, and green circles in the figure to the right).



2. All items (squares in the figure to the right) are associated to the "mean" that is closest to it based on geometric distance. This results in a set of *K* clusters, one cluster for every "mean".



3. The centroid of each of the clusters is then selected as a new "mean".



4. Steps 2 and 3 are repeated until some predefined conditions are met (number of steps, no significant change in the centroids, etc…)

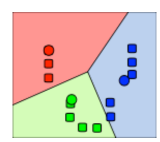


Figure 16. K-means clustering

One of the major difficulties with using *K*-means clustering is defining an initial value for *K*. It is tempting to set *K* to the number of classes being categorized (for example performing a binary classification by clustering into 2 groups). But there may be a richer, larger, more accurate set of latent clusters, based on unforeseen relationships between classification features, that together better constitute a desired class. To investigate if this intuition is supported by data a range of values for *K* was evaluated. Evaluation results were then used to characterize the sensitivity of the classification system to different values of *K*, and to suggest ranges of values for *K* that might produce optimal results for other data sets.

5.4. Feature And Classifier Evaluation

The final step in classifier design is evaluation. Classifier performance can be estimated by having the classifier categorize items from a gold standard. To establish if performance is acceptable the estimate can be compared to a predefined minimum performance, or it can be compared to the performance of some other classification system.

Gold standard and training data

Establishing a gold standard is often difficult. A gold standard can be created in a number of ways. If a classification system with a known level of performance exists, then its responses can be used as the gold standard. High levels of agreement with those responses would indicate that the evaluated classifier performances at a level approaching that of the gold standard system. An alternative is to use human judgments as a gold standard. This approach is illustrated in Figure 17 [162]. Yet another possibility is to use historical data as a gold standard. The use of historical data as a gold standard is only possible in situations where correct classifications will be easily available at a later date (e.g. stock performance, weather, and early diagnosis of disease).

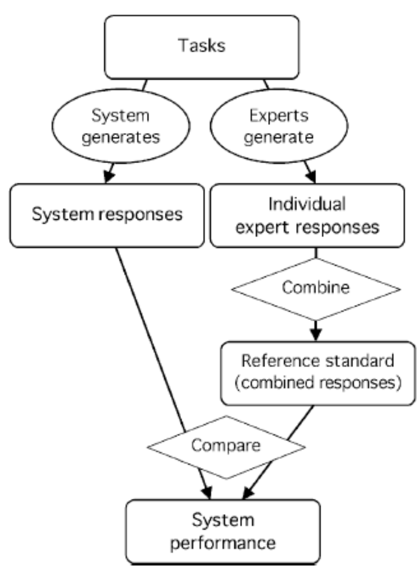


Figure 17: Estimating classifier performance using experts (Hripcsak 2002[162])

When evaluating systems that require training data, a decision must be made on where the training set will come from. One option is to create a training set that is independent of the gold standard. Another option is to use some of the gold standard items to train the classifier, and then use the remaining gold standard items for evaluating performance. One common method for using the gold standard for both training and evaluation is cross validation.

In cross validation the gold standard is repeatedly partitioned into training validation subsets. Each training subset is used to train the classifier, then the corresponding validation subset is used to estimate performance. The performance estimates over all the training/validation partitions are then aggregated (e.g. averaged).

One problem with this method is that for small data sets the estimated performance can be optimistic. Foley estimated that for binary classifications with a sample size per class of N, and l features, methods such as cross validation require that N/l be in excess of 3[163]. For example, if you have 10 features you need a minimum of 30 positive and 30 negative cases (i.e. minimum of 30 cases per class) to estimate performance. But even this may produce biased results, and so in general for smaller sample sizes cross validation is avoided in favor of independent training and validation sets.

Estimating performance

Once the gold standard is developed, and the way it will be used is established, metrics for estimating performance must be selected. The percentage of correct classifications, the accuracy, seems like a logical metric and is often used. But simple accuracy is very problematic. One drawback is that the percentage correct does not account for agreement due to chance. Another problem is that there may be different penalties and utilities associated with different types of correct and incorrect classifications.

One example where different penalties may favor one classifier over another is in the case of disease diagnosis and treatment. If a disease has a high mortality rate and the treatment for a disease has little or no negative side effects, incorrectly classifying a person as having that disease when they do not, referred to as a false positive or type I error, may be better (less costly or better for the potential health of the patient) than incorrectly classifying a person as not having the disease when they actually do have the disease, referred to as a false negative or type II error. Alternatively, if the treatment is very costly, or has major known side effects, or if going undiagnosed for some time does not pose an immediate health risk then type II errors may be better than type I errors.

Instead of simple percentage correct, binary classification metrics of sensitivity and specificity or the information retrieval metrics of precision, recall, and F-measure (the harmonic mean of precision and recall) are recommended[162].

		Correct Classifications	
		+	-
Predicted Classifications	+	True Positives (*tp*)	False Positives (*fp*) Type I errors
	-	False Negatives (*fn*) Type II errors	True Negatives (*tn*)

$$Precision = \frac{tp}{tp + fp} \qquad Specificity = \frac{tn}{tn + fp}$$

$$Recall, Sensitivity = \frac{tp}{tp + fn} \qquad F = 2 \bullet \frac{precision \bullet recall}{precision + recall}$$

Equation 5. Performance metrics

<u>How good is good enough</u>

What level of performance is good enough? As just stated this depends heavily on the task being performed. “Good” performance for a system that classifies news coverage may be completely unacceptable for medical diagnostic classifications. However there are some generally used definitions for different ranges of performance. For precision, recall and F-measure, results in excess of 0.8 are often defined as high quality, 0.6 – 0.8 are acceptable, less than 0.6 is typically viewed as poor[164,165].

Statistical significance of differences in performance between classifiers

When comparing the performance of different classification methods it is good practice to state if any observed difference is statistically significant. Statistical significance testing is a statistical hypothesis testing method where a test is used to measure if the difference between two aggregate statistics is greater than that allowed by chance to some level of certainty. When methods such as cross validation are used, performance estimates generated over the validation rounds can be averaged, then significance testing methods such as Student's *t*-test can be used to compare performance averages. But when multiple rounds of validation are not performed, and performance statistics cannot be aggregated or summarized (e.g. averaged) other tests such as the Wilcoxon test and the permutation test[††] must be used[166].

Tests such as the Wilcoxon and permutation tests use as their null hypothesis that the classifiers being compared produce the same responses and therefore do not differ significantly. Tests such as the Wilcoxon test, sign test and Student's *t*-test measure this by comparing the classifications of individual test cases (matched pairs). Of these, the Wilcoxon test has historically been the most popular among information retrieval professionals[166]. Unfortunately all three of these tests are limited in that they can only be used with the recall performance metric, but not with precision or F-measure. Two more flexible tests are the permutation test and bootstrapping. Both tests are more versatile because they can be used with any performance metric. However both methods are more computationally intensive than the aforementioned methods and have therefore only

[††] Sometimes refered to as the randomization test.

become widely used as computational power has become more readily available to researchers.

The permutation test was selected for use in this dissertation. The permutation test was chosen, and is recommended over the other methods, because unlike the Student's paired *t*-test, the sign test and the Wilcoxon test, it can be used with a variety of different performance metrics, unlike the bootstrap method it makes no assumptions regarding sampling from a population, and unlike Student's *t*-test it makes no assumptions regarding the distribution. The central idea behind the permutation test is that the null hypothesis, which states that the classifiers being compared are not different, can be demonstrated by showing that any response generated by one technique could just as likely have been produced by the other. This hypothesis is tested by having both classifiers categorize all test cases, then systematically reassigning the classifier responses to one of the two classifiers. The next step is to see how likely it is that these reassignments produced a difference in the performance that is as large or larger than the observed difference. The more likely it is that the systematic reassignment of classifier responses can result in a greater difference in performance, the more likely it is that the observed difference in performance was not significant. If the systematic reassignment of responses is not likely to result in a greater difference in performance, then the observed difference is more likely to be statistically significant.

One difficulty with using this method is that it is computationally expensive. For n test items there are 2^n different ways to assign the classifier responses. For a very small n one can try all test cases. But for any reasonably sized gold standard it becomes impractical to try all reassignments. For example a gold standard of size 20 generates over a million different reassignments.

For larger reassignment spaces one can sample a limited number of random permutations. While it is true that the more random permutations sampled, the more accurate the estimate, Smucker demonstrated that as few as 100K random permutations are sufficient[166].

Other problems with the use of this method are that it can sometimes produce spurious results, and it relies on the quality of the random-number generator used to generate permutations. To mitigate these problems Yeh suggested performing the permutation method twice and compared the results[167], and Smucker recommends that the random number generator used to perform the experiment be carefully selected[166].

Feature evaluation

It is often valuable to estimate the performance of individual classification features. One reason is that estimates of feature performance can assist in the proper selection of features during an iterative classifier design process (refer back to the discussion of classifier design in Section 5.1 page 50). Another reason, illustrated numerous times later in this dissertation, is that binary classifications are not always sufficient, especially for complex classification tasks or particularly difficult to classify items. When classification features are qualitatively relevant apart from the classifier they are used in, the reporting of their results may be of more consequence than the binary classification results. For both of these reasons it is important to identify which classification features are good discriminators of class.

Two tests that can be used for this task are Student's *t*-test and area under ROC (AUC). Student's *t*-test is useful when binary classifications are being performed and the average feature score for positive and negative items should differ significantly. Student's *t*-test is a widely accepted test of statistical significance for comparing sample means, and so provides a good indication of the value of a classification feature. However, while significant differences between feature mean scores can be a useful measure of feature performance, it is not always a sufficient test of class discrimination. For example, there may be a large difference between score means, but there may be a significant overlap of classes. To complement the *t*-test, the area under an ROC curve (AUC) is often used.

AUC is a common method for comparing classifiers. It reduces an ROC curve (a plot of a classifier's true positive rate versus its false positive rate) to a single scalar value representing expected performance. AUC is a good measure of the discrimination capacity of a classification feature primarily because it estimates the overlap of positive and negative items for that feature[168]. In addition, AUC has a number of other attractive properties. First, it is insensitive to class skew. If the proportion of positive and negative cases in a gold standard shifts (is skewed) in one direction or another, measures such as precision, and F-measure will be impacted. ROC curves, and AUC are based on true positive and true negative rates. If these rates are accurate then AUC should be consistent regardless of the number of positive or negative items in your gold standard. Second, AUC has the important statistical property of measuring a feature's ability to rank a randomly chosen positive test case higher than a randomly chosen negative test case. This is equivalent or closely related to other accepted classifier evaluation and significance testing metrics including the Wilcoxon test of ranks and the Gini coefficient. Third, AUC measures the actual overlap of cases, not based on an assumed normal distribution as in the case of *t*-test, but based on the actual observed classifications and scores. Lastly, AUC provides an easy, direct method for comparing individual features on their overall performance. For more information on AUC refer to Fawcett's introduction to ROC analysis[169].

5.5. Alternative classifiers: BoosTexter and SVM/LDA

The classifiers developed in this dissertation were compared to two other classification systems, the BoosTexter system[170], and a support vector machine (SVM) that uses latent Dirichlet allocation (LDA) topic distributions as features. These systems provided meaningful, "out of the box" alternatives that leveraged similar features to the classifiers developed for this project.

BoosTexter

The BoosTexter system is based on the technique of boosting[171], which is a general method for improving the performance of a supervised learning system. The basic thought underlying boosting is that a good learning rule, one that is strongly correlated to correct classifications, can be arrived at based on a set of much weaker rules. Boosting uses an iterative process of identifying the best rule not yet found for a classification task, and adding it to the overall rule set. This tends to force any additional rules that are subsequently added to focus on the harder training cases. Ultimately, once all the weak rules are combined, this system creates a classifier that is better than any of the individual weaker rules. The basic boosting method does not define the underlying classifier (the system that identifies or applies rules). This makes boosting a general procedure that can be used with whatever underlying classification system is appropriate to the classification task being performed.

BoosTexter is an implementation of boosting built specifically for text categorization. BoosTexter uses a simple one-level decision tree for its base learning system. The test at the root of the tree is whether a word or series of words appears in a text. Depending on the answer to this simple question BoosTexter predicts that the text belongs to a given category with some level of confidence.

SVM/LDA classifier

The second classifier is an SVM that uses LDA topic probabilities as features. The basic SVM algorithm assumes that the classes being distinguished are linearly separable. Figure 18 illustrates two solutions to a linear separability problem. SVMs find the solution that maximizes that margin between the training classes. In Figure 18 the solution in direction 2 is the best.

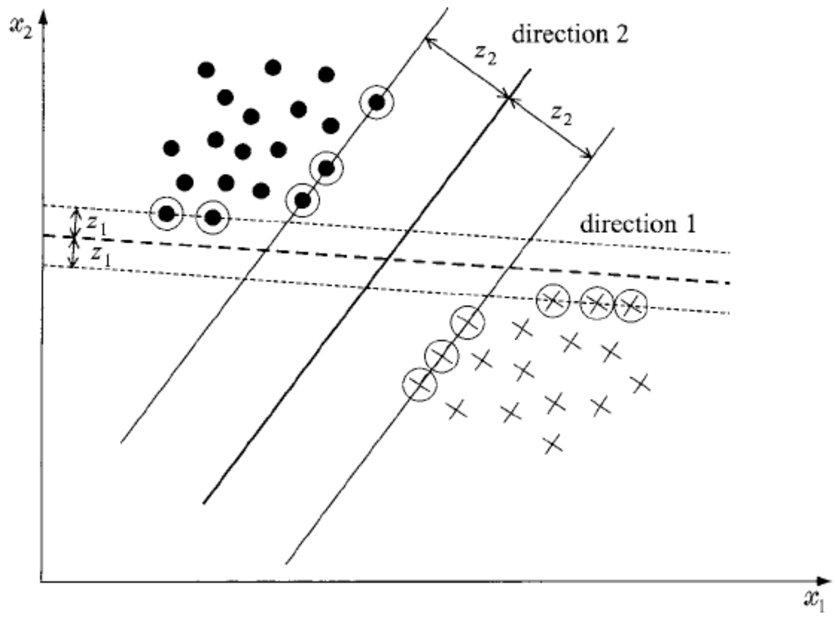


Figure 18. SVM: Example of 2 linear classifiers for the 2-class problem. (Theodoridis 2006[121])

As with many machine learning algorithms, SVMs are flexible in the type of input that can be used. One possible feature set for an SVM text classifier is the individual words that appear in a corpus. For example, a system could identify all unique terms in a corpus, then for each document in the corpus create a feature vector composed of the tf-idf score (refer to Equation 2) of each unique term. Once these feature vectors have been constructed an SVM can be trained for classification. Blei, in his original LDA paper, discussed how, for text classification tasks, the use of individual words can yield a classification rich feature set. However, when investigating the impact of reducing the information used in text classification by using LDA topics instead of the actual text words, he found that the lower order representation provided by LDA outperformed an SVM trained on actual text.

5.6. Summary

This chapter discussed computational and statistical methods relevant to this dissertation. In the first section, general classifier design was described. Classifier design follows a flexible iterative model where knowledge learned at every stage can lead to reevaluation of earlier design stages. In the second section I described the different classification features used in this dissertation. These include novel term and phrase detection, topic modeling, and sentiment scoring. In the third section I discussed methods for combining classification features into a single classification system. The *K*-means clustering algorithm was described in detail. In the fourth section discussed techniques for measuring the performance of classification and information retrieval systems. Specific topics that were discussed included the development and use of gold standards, information retrieval and classifier performance measures such as precision, recall and F-measure, and significance testing for deciding if differences in classifier performance are statistically significant. In the final section I described two classifiers, BoosTexter and SVM/LDA that were compared to the classifiers developed and evaluated in this dissertation. In the next chapter I will discuss a preliminary feasibility study that evaluated the use of one of the classification features used in this dissertation, *n*-gram language modeling, for use in distinguishing health news text from more general text.

Chapter 6. PRELIMINARY FEASIBILITY STUDY OF N-GRAM STATISTICAL LANGUAGE MODELING AND PERPLEXITY

Study first published in AMIA Annual Symposium Proceedings. 2009[131]

A preliminary study was conducted to investigate the feasibility of using *n*-gram language modeling and perplexity to differentiate news content related to the topic of obesity from more general news content. If the perplexity of *n*-gram language models (nLMs) of news content relative to models of obesity news increases as content becomes more general relative to obesity news coverage, then perplexity may be useful as a measure of relatedness of news coverage to news related to a health topic of interest (e.g. obesity). This result would support the first hypothesis of the second study of this dissertation, which states that there are patterns in the words and language of health news text that can be used to automatically classify health news by topic and framing.

6.1. Methods

The SalientNews RSS news aggregator[172] was used to create the corpus used in this research. I developed the SalientNews aggregation system as a tool for acquiring health news coverage from online RSS news sources, and for preparing the acquired news text for further study and analysis. SalientNews was used to gather all Reuters general and health news published online between January 1, 2008 and June 1, 2008. The Reuters health news corpus was filtered for obesity news by searching for news articles with titles or taglines containing any of the keywords; *"obesity"*, *"obese"*, *"diet*"*, *"weigh*"*, *"nutrition*"*, *"exercise*"* (where *"*"* represents a wildcard). Half of the resulting obesity articles were used to train an nLM and half was used as a test set in the evaluation. A Reuters general news corpus was created by including all non-health news published by Reuters. The English Gigaword[173] corpus, a comprehensive corpus of English newswire text, was sampled to create an even more general news corpus.

The Lingpipe[174] toolkit was used for text preprocessing (e.g. sentence chunking). The SRILM[175] toolkit was used to generate nLMs and calculate perplexities. Default modeling and perplexity settings were used (trigram models, Good-Turing discounting, Katz back-off smoothing). NLMs were generated from the first page of news article text published online.

Perplexities were calculated for all articles in the health and general news corpora, **and** for random samples of a size equal to the obesity news training data. Equivalently sized samples were included to ensure fair comparisons between corpora content. Perplexity where sentence boundaries are treated as words and perplexity where sentence boundaries are considered white space were calculated. Along with the perplexities, the number of out–of-vocabulary words (words in the test corpus that do not appear in the training corpus) was calculated.

6.2. Results

The general news corpus contained 25,104 news articles, the health corpus contained 2675 articles and the obesity corpus contained 289 articles (see Table 7). The obesity news language model was trained using a random sample of 145 obesity-related news articles.

The results (Table 8) show that perplexity increased as content became more general relative to the obesity news training data. This trend existed even when nLMs were generated using comparable corpus sizes. Perplexities were greater and increased more rapidly for more general content when sentence boundaries were considered. For corpora with the same number of news articles, the number of out-of-vocabulary words relative to the obesity news training data also increased as the content became more general.

Corpus	Articles	Sentences	Words
Reuters Obesity News (Reference)	144	1,551	41,143
Reuters Obesity News (Test)	144	1,630	42,308
Reuters Health News (small sample)	144	1,398	36,234
Reuters Health News	2,675	32,206	730,285
Reuters General News (small sample)	144	1,398	34,614
Reuters General News	25,027	233,778	5,854,300
Gigaword (small sample)	121	2,681	56,233
Gigaword	19,825	394,849	8,460,935

Table 7. Perplexity Study Corpora

Corpus	Perplexity (w/sent. bounds)	Out of Vocab Words
Reuters Obesity News	187 (236)	5,335
Reuters Health News (small sample)	278 (366)	7,431
Reuters Health News	275 (362)	135,532
Reuters General News (small sample)	378 (542)	11,614
Reuters General News	372 (531)	1,966,504
Gigaword (small sample)	382 (591)	19,604
Gigaword	387 (594)	2,966,079

Table 8. Perplexity Study Results

6.3. Summary

The results of this preliminary study indicate that the perplexity of nLMs can be used to differentiate obesity news from more general news text. This suggests that perplexity can be used as a classification feature in the development of an automated health news topic classifier.

In the next chapter I describe the methods used to implement and evaluate the health news classifiers developed in this dissertation. These classifiers used nLMs, along with other features, to classify health news on its relevance to the topic of obesity and on its health news framing.

Chapter 7. HEALTH NEWS CLASSIFIER STUDY: METHODS

In this chapter I describe a study in which software was developed to classify health news coverage on its relevance to the topic of obesity and on its news framing. The classifiers, referred to in this chapter as feature classifiers, or F-classifiers, were evaluated by estimating performance against expert human responses, and by comparing performance to two machine-learning classifiers and (in the case of the obesity news classifier) a simple keyword based classifier.

The research questions and hypotheses for this study were:

RQ1: What level of performance can be achieved for classifying health news on relevance to obesity and on news framing using unsupervised and semi-supervised classification methods?

RQ2: What classification features perform best for these tasks?

H1: There are patterns in the words and language of health news text that can be used to automatically classify health news by topic and framing.

H2: Classifying based on a news article's framing is a more difficult task than classifying based on broad health topics such as obesity.

H3: Acceptable performance can be achieved with very limited training data.

The following sections describe in detail the methods used to implement and evaluate the classifiers. For clarification on the definition of computational concepts, methods and terms, please refer back to the appropriate section in Chapter 5, and to the glossary of information retrieval and classification terms on page 49 of Chapter 5.

7.1. Corpus and gold standard

The Reuters health news corpus used in the preliminary study described in the previous chapter was also used in this study. That corpus was comprised of news articles published online by Reuters in 2008. Reuters categorized by news articles in the corpus as health related (*n*=3227). News articles were collected using the SalientNews RSS aggregator[172].

Two gold standards were created for this study. The first was a gold standard of news articles classified on their relevance to the topic of obesity. The second was a gold standard of news articles classified on their overall health news framing. These gold standards were developed using the responses of human experts.

<u>Obesity news gold standard</u>

For the obesity news gold standard two expert human coders categorized 130 news articles randomly selected from the Reuter's health news corpus. The use of two coders is consistent with numerous health news content analyses [2,9,32,38,44,65,114,176,177]. The number of articles coded was selected to strike a balance between creating as large a reference as possible while keeping the workload relatively low (~ 2 hour). Coders rated news articles on a four-point scale based on relevance to the topic of obesity; 0 = *not at all related*, 1 = *remotely related*, 2 = *moderately related* or 3 = *very much related* (see APPENDIX B on page 187 for illustrations of the web interface used to collect responses). A balanced distribution of news articles, ranging from news articles that were completely unrelated to obesity, to news articles very much related to obesity, was sought. It was expected that a relatively small percentage of all health news articles would be related to the topic of obesity. To ensure that a balanced gold standard was created, a subset of gold standard news articles (70%) were randomly selected from the news articles that passed the obesity news keyword filter used in the preliminary study described in the last chapter (see page 77). Two thirds of the remaining gold standard news articles (20% of the total) were randomly selected from the rest of the health news corpus, and the remainder (10% of the total) were randomly selected from general news.

Individuals with experience in health communications, health journalism or health news research were recruited for this task. Coders were recruited using word of mouth from the Columbia University community and through peers and colleagues.

The R-project[178] implementation of Krippendorff's alpha[39] was used to calculate interrater agreement. Krippendorf's alpha was selected because of its facility with calculating agreement based on ranked human responses.

A binary classification of relevance to obesity was derived from the coder responses. If both coders categorized a news article using one of the three codes that indicate relevance to obesity then the news article received a positive gold standard classification. If both coders classified the article as not related to obesity then the article received a negative classification. For cases where a binary classification was not possible using this strategy, a third expert coder was used to provide a final categorization. Coders were also able to enter comments regarding their response and the news article being coded.

Apart from calculating agreement and the balance of positive and negative cases, the obesity gold standard classifications were not examined in detail until after the classifiers were developed and evaluated.

Health news framing gold standard

Three coders were recruited to develop the health news framing gold standard. People with experience in health communications, health journalism or health news research were recruited as coders. Coders were recruited by word of mouth from the Columbia University community and through peers and colleagues. Coders were asked to give binary classifications on the framing of 80 health news articles randomly selected from the obesity news gold standard (see APPENDIX B on page 187 for illustrations of the web interface used to collect expert responses). The 9 health news framings that coders were allowed to apply are listed in Table 9.

Framing Group	**Health news framing**	**Description (This framing discusses…)**
Cause, Blame, Responsibility	Genetic/biological causes[23,65]	biological causes or contributors to health
	Environmental[23,65,179]	risks imposed upon a person by their environment
	Economic issues[43]	potential or realized impact of economic issues due to exposure, treatment or health consequences
	Personal Responsibility[23,25,42,65]	an individual’s role and responsibility in maintaining health
	Culture / Community[42,43,65]	the role culture and community plays in risk, disease prevention, treatment and health outcome
	Lifestyle[13,15,18]	the role lifestyle plays on disease prevention, treatment and health outcome
Intervention and Treatment	Research advances[65]	research results that can be used to improve the health of individuals or groups of people
	Medical interventions[36]	the impact of new and existing medical interventions (e.g. surgery, drugs)
	Psychosocial interventions[65]	the impact of interventions that try to modify both psychological and social aspects of an individual's behavior

Table 9: Health news framings

For each news article experts were asked to select one or more health news framings, or indicate that none of the available framings were appropriate. Coders also had the option of enter comments about each news article and their response. Classifiers were to be developed only for the health news framings with the best (most even) distributions of positive and negative cases.

Cohen's kappa[180] as implemented in R-project[178] was used to calculate interrater reliability. The majority's opinion of the classifications for each news article was used as the gold standard classification. In the case of low interrater agreement additional rounds of coding would be performed.

Like the obesity gold standard, apart from assessing the numbers of positively and negatively coded cases, and measuring coder agreement, the framing gold standard classifications were not reviewed until after the classifiers were developed and evaluated.

7.2. Classifier Implementation

Classifier inputs and outputs

The F-classifiers developed in this study took as inputs the entire Reuter's health news corpus, and news articles from the health corpus designated as training data. The primary F-classifier output was the binary classifications of all tested news articles. In addition all intermediate classifier feature data was also outputted for evaluation.

The news articles used as training data for the classifiers were generated from the responses of non-expert coders who manually reviewed a random selection of news articles from the health news corpus. These non-expert coders were asked to code news articles based on relevance to the topic of obesity, and for the framing of the health news coverage (similar to the coding performed by experts to generate the framing and obesity news gold standards). Non-expert coders were recruited using word of mouth from the Columbia University community and through peers and colleagues. Individuals with no special experience in health communications or health news analysis were recruited for this task.

The news articles selected by the non-experts were reviewed and refined by adding and removing news articles as needed to achieve the final training sets. The planned size of the F-classifier training sets was approximately 40. This size was selected because it is very limited (smaller than a typical machine learning training set), and convenient (could be easily, quickly, manually created by an individual researcher that needs to retrieve or classify health news coverage). Each training set was composed of 50% positive cases, and 50% negative cases.

Preprocessing

The following steps were taken to prepare news text for analysis:

1. **Chunking**: News articles were broken up into sentences.
2. **Stop-list**: A stop-list was used to remove irrelevant high-frequency terms.
3. **Semantic labeling**: For the obesity and framing news classifiers semantic class labels replaced certain terms and phrases. Examples included proper names, place names, company names, organization and institution names, dates and quantities. In addition, for the framing F-classifiers health conditions and treatments were also replaced with generic semantic class labels. This additional step was taken so that framing classifications would not be dependant on health conditions or treatments.
4. **Stemming**: All remaining words (those not removed by the stop-list or replaced by a semantic label) were stemmed using the Porter Stemming algorithm[181].

Text preprocessing was implemented using a combination of PHP scripts, Bash shell scripts, and Java programs. The Lingpipe[174] toolkit was used for chunking and stop-list filtering. OpenCalais[182] and extended regular expressions were used to replace words and phrases with semantic class names.

Classifier Features

The feature set used by all F-classifiers included; the length of news articles (in characters), novel term/phrase (NTP) scores, *n*-gram language model (nLM) perplexity, and latent Dirichlet allocation (LDA) scores. Additional features used by the framing F-classifiers were text readability, affect scores, and the frequency of certain word types. These additional features were used for framing classification because of the potential increased complexity and breadth of the framing classification task.

Unigrams, bigrams and trigrams that had the highest *z*-scores in the classifier's training set relative to the overall health corpus were considered to be valid NTPs. This is similar to the SIPS method used by Amazon.com (see Section 5.2 on page 54). For each classifier a minimum of approximately 10 NTPs was desired. The achieve this minimum a minimum *z*-score of 15 was used for unigrams and bigrams, and a minimum score of 10 was used for trigrams. These minimum *z*-scores were arrived at by testing a range of *z*-score values. The LingPipe[174] toolkit was used to calculate NTP *z*-scores.

In the obesity news classifier the NTP feature value was based on the location of that NTP in a tested news article. If an NTP appeared in a news article then the feature score was set to the location of the first occurrence of the NTP, normalized by the size of the news article to a value between zero and one. For example, if an NTP appeared at the 10^{th} character in a news article containing 100 characters, that news article would receive a score of .1 for that NTP. If an NTP did not appear in a news article then that news article received a score of 100 for that NTP. This was done to drastically separate the scores of news articles where NTPs appear from articles where NTPs did not appear.

For the framing classifiers NTPs were scored using the tf-idf measure (see Section 5.2 on page 54). NTPs were scored differently in the framing classifiers because initial testing revealed that NTPs relevant to classification based on framing were more likely to be scattered throughout the news article and would be better measured using the tf-idf calculation, whereas NTPs that indicate that a news articles is relevant to a specific health topic are more likely to be prominently placed early in that news article.

To compute *n*-gram language model (nLM) perplexity, a reference nLM was constructed using positive training cases. The perplexities of the nLMs of all news articles in the health news corpus relative to positive training case nLM were then calculated. The SRILM[175] toolkit was used to generate nLMs and calculate perplexities. Default modeling and perplexity settings were used (trigram models, Good-Turing discounting, Katz back-off smoothing).

LDA was used in all classifiers. This was done specifically to combat the potential problems of relying too heavily on keywords. LDA scores were calculated by first generating an LDA index for the entire health news corpus, then querying using a pseudo-document generated by concatenating positive training cases. The LDA scores resulting from querying the index were used as the LDA score features. Various numbers of LDA topics were tested during implementation. The best performing topic counts (values of *K*) for the obesity and framing classifiers are reported. The GibbsLDA++ software package[183,184] was used to generate the LDA index. Default LDA indexing settings of 0.5 for alpha, and 0.1 for beta, were used. Kullback-Leiber divergence[130] was using to calculate LDA scores.

Text readability was calculated using the Flesch Reading Ease Score (FRES) [145]. Flesch Reading Ease Score (FRES) was selected because unlike most other measures that attempt to approximate the US grade level needed to comprehend the text, FRES provides a score between 0 (best understood by university graduates) and approximately 120 (easily understood by an average child of 11 years). This wider range of scores facilitated implementation, and the larger variance that resulted from this range facilitated statistical analysis.

Whissell's Dictionary of Affect in Language (DAL) [154] and the Linguistic Frequency and Word Count dictionary (LIWC) [156] were used to create features that measured the conceptual and emotional content of news text. The means of the three DAL dimensions *pleasantness*, *activation* and *imagery,* across all words in a news article was used to create 3 separate features. All LIWC word subcategories under the categories of *Social Processes, Affective Processes, Cognitive Processes, Perceptual Processes and Biological Processes* were also used as features. LIWC features were calculated by summing the number of words from a category that appear in a news article, divided by the number of words in the news article. Refer to Appendix C on page 195 for examples from these dictionaries.

<u>Classification</u>

Based on the described features, a news article's feature vector was defined as:

Obesity classifier feature vector = $[l, P, L, N_1, \ldots, N_n]$

Framing classifier feature vector = $[l, P, L, N_1, \ldots, N_n, R, A, W_1, \ldots, W_n]$

Where:

Framing Classifier features	***Obesity Classifier features***	l =	Length of the news article
		P =	nLM perplexity value
		L =	LDA score
		N_i =	Score of NTP *i.* For the obesity classifier this is the normalized location of the NTP, or 100 if the NTP is not present in the text. For the framing classifiers an NTPs feature score was its tf-idf score.
		R =	FRES readability score
		A =	Affect scores
		W_i =	Word type frequency

Equation 6. Classifier feature vector

News article feature vectors were generated, and then a semi-supervised clustering analysis was used to group and classify all health news articles. Clustering analysis was done using the *K*-means clustering algorithm[160]. A cluster's news articles were labeled based on the number of positive and negative training items in that cluster. If no training articles were in a cluster then all news articles in that cluster were labeled negative. If there were positive training articles but no negative ones then the cluster was labeled positive. If there were negative training articles but no positive ones then the cluster was labeled negative. If there were positive and negative training articles then the cluster received the label of the training article closest to the geometric center of the cluster. Figure 19 illustrates these rules.

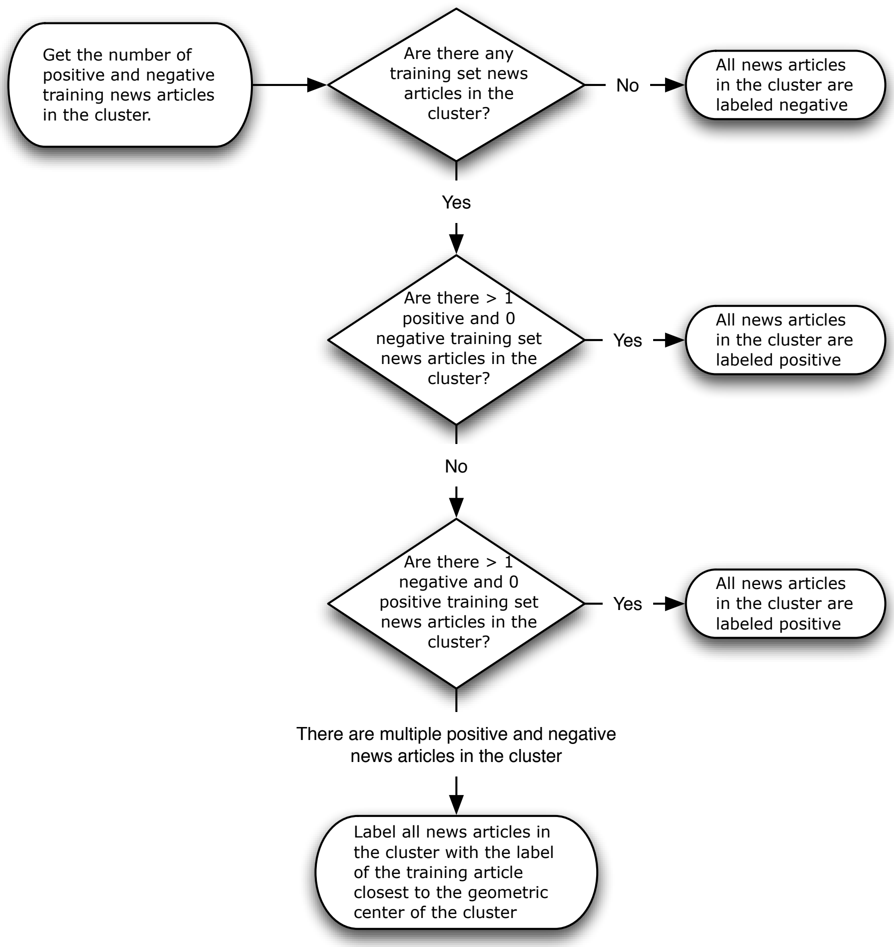


Figure 19. Cluster labeling flow chart

For evaluation the value of *K*, the number of clusters generated by the *K*-means algorithm, was systematically tested for all values between 3 and 35. The Weka data mining software package[185] was used to perform the clustering operation.

7.3. Evaluation

Classifier evaluation

Classifier performance was estimated by calculating precision, recall, and F-measure against the gold standards. Positive findings were defined as achieving precision, recall and F-measure results in excess of 0.8 . The F-classifiers developed for this dissertation were also compared to the BoosTexter system, a support vector machine classifier (SVM) that uses LDA topic distribution features, and (in the case of the obesity relevance classification task) an obesity keyword filter. The keyword filter categorized news articles as relevant to the topic of obesity if any of the following keywords[‡‡] appears in the title of the news article; *obesity*, *obese*, *diet, weigh*, nutrition*,* or *exercise**. This keyword query was defined using queries from other studies of obesity related health news [24,38,40,131].

BoosTexter[170] and the SVM/LDA classifiers were trained on the unprocessed text[§§] of the news articles in the training sets. For the SVM/LDA classifier the only parameter varied during testing was the range of topic sizes (from 50 to 225). BoosTexter allows for the modification of the number of training rounds, the *n*-gram method (*n*-grams or sparse *n*-grams), and the window size (distance between the first and last terms of *n*-gram/sparse *n*-gram). For the classifier evaluation, BoosTexter window size was fixed to 3 (trigrams), the number of training rounds was varied by increasing orders of magnitude between 10 and 100000, and both distance methods (*n*-gram and sparse *n*-gram) were tried in turn. For all F-classifiers the size of the training set was systematically increased to identify the optimal classifier. The GibbsLDA++ software package[183,184] was used to generate the LDA index. The Weka data mining software package[185] was used to generate the SVM models.

For the framing F-classifiers, in addition to overall performance, their performance against the subset of news articles in the framing gold standard that were positive cases in the obesity news gold standard was estimated.

The significance of differences in performance of all classifiers was tested using the permutation test[***] (see Section 5.4 page 69 for a discussion of the permutation test). The permutation test was run twice for each test of significance, with 1 million randomly

[‡‡] In this query an astrisk represents a wildcard of 1 or more letters.
[§§] Not run through the preprocessing steps described on page 85.
[***] The permutation test software developed for this dissertation is publically available at http://blog.ledona.net/techblog/2010/11/20/permutation-test/

selected permutations per run as recommended by Yeh[167]. If both permutation test runs produced the same qualitative answer (i.e. a significant difference, or no significant difference) then the quantitative result of the first test was reported.

Feature evaluation

Classifier feature performance was also estimated. The NTPs of the best performing F-classifiers, and the average scores of positive and negative cases in all gold standards for each NTP, news article length, LDA, nLM, DAL, LIWC, and readability score were reported. DAL, LIWC, news article length, and readability are static features that are not dependant on the training data. The performance of these features is reported along with descriptive data for the gold standards. The remaining features are based on training data, and are reported with the results for the optimal F-classifiers.

Each feature's ability to discriminate between positive and negative gold standard cases was tested using Student's *t*-test and AUC. AUC scores range from 0.0 to 1.0. A score of 0.5 typically indicates that a feature does not perform better than a random classifier/discriminator. Scores greater than 0.5 indicate that the feature performs better than chance, lower than 0.5 means worse than chance. A score of 1.0 is a perfect classifier. A score of 0.0 is a perfect inverse classifier (always wrong). AUC was calculated using the ROCR package in R-project[178]. Refer to Table 10 for interpretations of different AUC ranges.

Feature Performance	**AUC score**
Excellent	0.85 – 1.0
Very good	0.7 – 0.85
Good	0.6 – 0.7
Poor	0.5 – 0.6

Table 10. Student's t-test and AUC performance ranges

Student's *t*-test was also used to calculate each feature's capacity to discriminate class based on differences in sample means of positive and negative cases[168]. The use of Student's *t*-test can be problematic for measuring feature performance in this manner because of *p*-value sensitivity to sample size and variance. But because *t*-test is often used for this purpose, it is included here. Student's *t*-test was calculated using Microsoft Excel 2008 and 2011.

Chapter 8. HEALTH NEWS CLASSIFIER STUDY: RESULTS

This chapter contains the results of the obesity news and health news framing classification studies. In Section 8.1 I describe the results of the development of the obesity news and health news framing gold standards. Sections 8.2, 8.3, and 8.4 contain the evaluation results for the obesity news classifier, and the 2 framing classifiers (lifestyle framing and biological/genetic causation framing). In Section 8.5 I discussion the study results.

8.1. Gold standards

Obesity news gold standard

The level of agreement between the 2 coders (C1 and C2) for the obesity news gold standard was α=0.751. A third coder (C3) was used to resolve all disagreements between C1 and C2 (*n*=21). When reduced to binary classifications, 71 news articles were categorized as relevant to the topic of obesity, and the remaining 59 news articles were not relevant to obesity.

In all cases where C1 and C2 disagreed (*n*=21), C1 coded the news article as *not related to obesity*, while coder C2 coded 11 cases as *remotely related to obesity*, 9 cases as *somewhat related to obesity* and 1 case as *very related to obesity*. An examination of news article text, and coder comments[*] revealed that most of the disagreements turned on how to categorize news articles that discussed topics that were objectively related to obesity, but where obesity was not directly discussed. One example of this was a news article discussing the benefits of exercise to patients suffering from cancer-related fatigue[†], another example

---

[*] This assessment was performed after all classifier development and evaluation was complete.
[†] Full text on page 199

example was a news article describing research that links nonfat milk to prostate cancer*. Both discuss topics related to obesity, but the text in the news articles did not explicitly discuss obesity. Coder C3, when resolving the disagreement between the other coders, expressed this by commenting, *"though you should drink non-fat milk if you're obese, still not about obesity"*.

Article length was the only static feature (a feature not based on training data) used by the obesity news F-classifier. This feature was a very good discriminator of positive and negative class for obesity relevance based both on the AUC score and on p-value (1904.9 vs. 1524.2;p-value<0.001;AUC=0.804).

Framing gold standards

Two rounds of coding were needed to generate the framing gold standard. The first round revealed that two news framings, lifestyle and genetic/biological causation, had the best distributions of positive and negative cases. However, the level of agreement between coders in the first round was low (κ of 0.515 and 0.507) so a second round of coding was conducted to increase consensus (see Table 11).

During the second round coders were asked to review, and potentially update their responses to 35 news articles where unanimous agreement was not achieved in the first round. To ensure that coders did not feel pressured to change their original responses they were not told when their original responses were in the majority or minority, and they were instructed not change their response unless they thought they answered incorrectly in the first round. After the second round of coding, interrater agreement for the lifestyle and genetic/biological issues frames improved to acceptable levels (κ = 0.711 for genetic/biological causation, κ = 0.842 for lifestyle).

* Full text on page 200

Health news frame	1st round of coding Positives	κ	2nd round of coding Positives	κ
Culture / Community	13	0.714		
Economic Issues	2	0.942		
Environmental Risk	3	0.889		
Genetic/Biological Issues	44	0.515	40	0.711
Lifestyle	36	0.507	40	0.842
Medical interventions	19	0.62		
Personal Responsibility	11	0.711		
Psychosocial interventions	5	0.845		

Table 11. Health news framing gold standard

Lifestyle gold standard classifications and static features

Many of the lifestyle framed news articles discussed behavioral interventions such as diet and exercise. For example, news articles such as one titled "*Exercise and limited TV time may keep kids trim*"[*], and another titled "*Nature tops nurture in childhood obesity: study*"[†], were unanimously coded as lifestyle framed news articles. News articles that used more scientific, medical, or technical language, and news articles that delved deeply into biological processes, disease symptoms, medical interventions, doctor/patient interactions and legislative or business related news coverage, tended to not be categorized as lifestyle news articles. Examples of news articles unanimously coded as not having a lifestyle health news framing included one titled "*Hormone discovery may help combat diabetes: study*"[‡], and another titled "*Unilever says new milkshake helps control appetite*"[§].

Table 12 (on page 97) contains a comparison of average static feature scores between lifestyle and non-lifestyle news articles. As with the obesity news gold standard, news articles coded as having a lifestyle framing were significantly longer than other news

[*] Full text on page 208
[†] Full text on page 210
[‡] Full text on page 211
[§] Full text on page 213

articles. All three DAL dimensions were good discriminators of lifestyle framing based on both *t*-test significance and AUC. LIWC categories identified as good discriminators of lifestyle framing by AUC or *t*-test were the *social processes* category, its *friends* and *humans* subcategories, the *certainty* and *inclusive* subcategories of *cognitive processes*, and the *biological processes* category of *ingestion*.

The 3 coders (Ca, Cb, Cc) tended to use different strategies to categorize the more difficult to classify news articles. Based on the comments left by the coders during the 2 rounds of coding, coder Ca tended to code news articles based on the broad concepts discussed throughout the news article text. Coder Cb tended to weigh the various specific topics and concepts that appeared within a news article's text, and coded for the more prominent of those found. Cc tended to code for lifestyle framing based on the existence of any mention of a concept directly related to lifestyle framing. For example, near the end of a news article describing research on a decreased risk of colon cancer for women that use birth control pills[*], there is text describing the possible impact of weight loss, physical activity and lifestyle changes. Cc coded this article as lifestyle, and commented that *"Physically active and weight loss = lifestyle"*. Ca and Cb did not code this as a lifestyle news article, likely basing their response on the entirety of the news article text.

---

[*] Full text on page 203

	Mean scores of GS articles			
	Positive	**Negative**	***p*-value**	***AUC***
Article length	2018.975	1701.000	**0.005**	**0.678**
FRES	51.355	53.010	0.417	0.533
DAL dimensions				
Activation	1.371	1.319	**0.006**	**0.669**
Imagery	1.275	1.211	**0.001**	**0.703**
Pleasantness	1.463	1.402	**0.002**	**0.733**
LIWC categories and subcategories				
Social Processes	0.081	0.072	0.131	**0.601**
Family	0.002	0.000	0.068	0.597
Friend	0.003	0.002	0.080	**0.632**
Humans	0.023	0.017	**0.025**	**0.640**
Affective processes	0.036	0.034	0.436	0.563
Negative emotions	0.021	0.021	0.997	0.508
Anger	0.015	0.012	0.187	0.549
Anxiety	0.003	0.002	0.455	0.549
Sad	0.007	0.007	0.943	0.551
Positive emotions	0.007	0.007	0.654	0.577
Cognitive processes	0.133	0.136	0.597	0.521
Causation	0.017	0.022	0.052	0.598
Certainty	0.006	0.004	0.077	**0.631**
Discrepancy	0.007	0.009	0.192	0.548
Exclusive	0.015	0.017	0.295	0.589
Inclusive	0.045	0.040	0.111	**0.639**
Inhibition	0.008	0.008	0.732	0.547
Insight	0.021	0.023	0.383	0.565
Tentative	0.022	0.023	0.742	0.517
Perception	0.021	0.021	0.826	0.536
Feel	0.007	0.008	0.712	0.500
Hear	0.007	0.008	0.532	0.525
See	0.004	0.003	0.706	0.528
Biological processes	0.098	0.086	0.093	0.599
Body	0.020	0.020	0.955	0.516
Health	0.062	0.057	0.403	0.556
Ingestion	0.033	0.020	**0.013**	**0.661**
Sexual	0.003	0.003	0.886	0.526

Table 12. Static feature scores for lifestyle framing gold standard

A more difficult example that illustrates the difference between coding styles can be observed in the coding of a news article that discussed how childhood obesity rates in France were leveling off perhaps due to healthier diet programs[*]. Ca coded this news article as not having a lifestyle framing, and commenting that, *"this article is really about the success of policy changes/programs in France (and not so much about the changes in children's lifestyle)"*. Cb disagreed, coded this as a lifestyle news article, and commented that *"the article focuses on policy /environmental changes that influences lifestyle choices"*.

<u>Biological/genetic causation gold standard classifications and static features</u>

News articles that were coded positively for biological/genetic causation framing typically discussed biological or genetic associations or processes. Examples included a news article discussing obesity's negative impact on male fertility[†], and another discussing how low thyroid activity has been linked to weight gain[‡]. News articles that primarily discussed issues that were not genetic or biological in nature, or did not discuss in some way proven or possible causal processes generally were not coded positively for biological/genetic causation framing. Examples of news articles not coded for biological/genetic causation included one discussing the health benefits of a strict Mediterranean diet[§], and another discussing the popularity of multivitamins among teens[**].

Table 13 lists a comparison of average static feature scores between positive and negative cases from the biological/genetic causation framing gold standard.

---

[*] Full text on page 204.
[†] Full text on page 214
[‡] Full text on page 216
[§] Full text on page 217
[**] Full text on page 218

	Mean scores of GS articles			
	Positive	**Negative**	***p*-value**	***AUC***
Article length	1867.575	1852.400	0.895	0.581
FRES	51.390	52.975	0.437	0.514
DAL dimensions				
Activation	1.331	1.359	0.152	0.581
Imagery	1.233	1.252	0.356	0.518
Pleasantness	1.416	1.449	0.093	0.564
LIWC categories and subcategories				
Social Processes	0.074	0.079	0.411	0.547
Family	0.001	0.002	0.469	0.579
Friend	0.003	0.002	0.184	0.593
Humans	0.023	0.018	0.077	**0.626**
Affective processes	0.034	0.036	0.649	**0.677**
Negative emotions	0.022	0.020	0.433	**0.634**
Anger	0.002	0.003	0.584	0.528
Anxiety	0.008	0.005	**0.047**	**0.614**
Sad	0.007	0.007	0.927	**0.636**
Positive emotions	0.012	0.016	0.061	0.562
Cognitive processes	0.134	0.134	0.946	0.523
Causation	0.021	0.018	0.292	**0.712**
Certainty	0.005	0.005	0.744	**0.603**
Discrepancy	0.008	0.009	0.314	0.511
Exclusive	0.015	0.017	0.261	0.560
Inclusive	0.040	0.045	0.167	**0.616**
Inhibition	0.008	0.009	0.563	0.557
Insight	0.024	0.020	0.069	0.528
Tentative	0.024	0.022	0.343	0.528
Perception	0.020	0.022	0.660	0.553
Feel	0.008	0.006	0.272	0.519
Hear	0.006	0.010	**0.047**	0.599
See	0.004	0.004	0.941	0.523
Biological processes	0.102	0.082	**0.005**	**0.648**
Body	0.026	0.013	**0.001**	**0.661**
Health	0.064	0.054	0.091	0.576
Ingestion	0.028	0.025	0.527	**0.610**
Sexual	0.003	0.003	0.697	0.501

Table 13. Static feature scores for biological/genetic framing gold standard

Some static features that were good indicators of lifestyle health news framing, for example news article length and the 3 DAL dimensions, were not good indicators of biological/genetic causation framing. But there were numerous LIWC features that did prove to be good indicators. Examples included the *affective processes* category, the *negative emotions* category, and its *anxiety* and *sad* subcategories, and the *biological processes* category, and its *body* and *ingestion* subcategories.

Disagreements between coders for the biological/genetic causation framing gold standard were similar to disagreements in the coding of lifestyle framing. When there was disagreement it appeared to be due to coder Ca coding based on the overall news story, Cb coding based on the prominence of the different concepts discussed within a news article, and Cc coding on if biological/genetic causation concepts appeared anywhere within a news article. For example the coders disagreed on the coding of a news article that discussed the impact that a lack of sleep has on children[*]. Ca coded this negatively, and commented that the article was discussing the *"relation of sleep (i.e. behavior/lifestyle) to obesity"*. Cb and Cc coded this positively for biological/genetic causation framing, both citing text that occurs in the second half of the news article. Cb commented, *"This article [discusses] the biology and correlations between sleep patterns and weight. The conclusion leads to recommendation for lifestyle change thus I would select both B/C and Lifestyle"*. Coder Cc commented quite simply, *"Hormone levels (biological issue) that affect hunger"*.

Disagreement also arose based on how the 3 coders interpreted the relevance of various concepts to biological/genetic causation. One example was a news article discussing the prevalence of pelvic floor disorders among woman[†]. Ca coded that this was not framed as a biological/genetic causation story, commenting that the news article discussed *"prevalence of pelvic floor disorders without delving into too much biological detail*". Cb and Cc disagreed, and coded this news article positively for biological/genetic causation framing. Cb, in the second round of coding, commented that, *"assuming discussion of physical symptoms, physiology, and biological issues are included in the category of*

---

[*] Full text on page 241

[†] Full text on page 207.

"genetic/Bilogical", I'm going to keep my answer". Cc commented, *"being a woman, seems like a genetic issue*".

Readability

Readability as measured by FRES was not a good indicator of either lifestyle or biological/genetic causation framing. However, it should be noted that the average readability scores measured in this study (FRES between 50 and 53) are indicative of text that is fairly difficult to read. Texts of this readability level are not recommended for material targeted at average American readers[145]. This finding is consistent with what other researchers have found in general audience health news coverage [74,81,88].

8.2. Obesity news classification results

Overall performance of obesity classifiers

The obesity news F-classifier (implemented for this dissertation) out performed all the evaluated classifiers, and significantly outperformed both the keyword filter and the SVM/LDA classifier. In its optimal configuration the F-classifier had an F-measure of 0.923 (see Table 14). The F-classifier's optimal performance was achieved using a training set of 24 news articles, an LDA index of 100 topics, and 10 clusters. BoosTexter was second best in performance with an F-measure of 0.894, accomplished over all tested numbers of training rounds, standard and sparse trigrams, and training set sizes of 38 and 40. The keyword filter followed with an F-measure of 0.852. The SVM/LDA had the lowest performance with an F-measure of 0.797. The SVM/LDA classifier achieved this using a query size of 40, and an LDA index of 50 topics.

Obesity news classifier	**Precision**	**Recall**	**F**
F-classifier	0.917	0.930	0.923
BoosTexter	0.900	0.887	0.894
Keywords	0.758	0.972	0.852
SVM/LDA	0.792	0.803	0.797

Classifiers	**Measure**	**Difference**
F-classifier vs. Keyword filter	Precision	**0.158 (*p*<0.001)**
	Recall	-0.042 (*p*=0.375)
	F-measure	**0.071 (*p*=0.015)**
F-classifier vs. BoosTexter	Precision	0.017 (*p*=0.126)
	Recall	0.042 (*p*=0.250)
	F-measure	0.029 (*p*=0.126)
F-classifier vs. SVM/LDA	Precision	**0.125 (*p*=0.018)**
	Recall	0.127 (*p*=0.063)
	F-measure	**0.126 (*p*=0.004)**
Keyword filter vs. BoosTexter	Precision	**-0.142 (*p*=0.001)**
	Recall	0.085 (*p*=0.070)
	F-measure	-0.042 (*p*=0.195)
Keyword vs. SVM/LDA	Precision	-0.033 (*p*=0.458)
	Recall	**0.169 (*p*=0.002)**
	F-measure	0.055 (*p*=0.141)
BoosTexter vs. SVM/LDA	Precision	**0.108 (*p*=0.047)**
	Recall	0.085 (*p*=0.264)
	F-measure	**0.096 (*p*=0.035)**

Table 14. Obesity news classification results and comparisons

Figure 20 illustrates the overall trends in performance of the F-classifier, SVM/LDA classifier and BoosTexter as training set size was increased from 6 to 40. For most training set sizes the F-classifier's performance was clearly better than the other classifiers. All 3 classifiers begin with performance between 0.6 and 0.7. The performance of the F-classifier and BoosTexter increased dramatically until the training set size reached 10, and then remained essentially level. BoosTexter performance only increased appreciably again at training set size of 38. At that point its performance peaks, almost matching the maximum performance of the F-classifier. The SVM/LDA classifier showed constant improvement in performance throughout, but never matched the performance of the other classifiers.

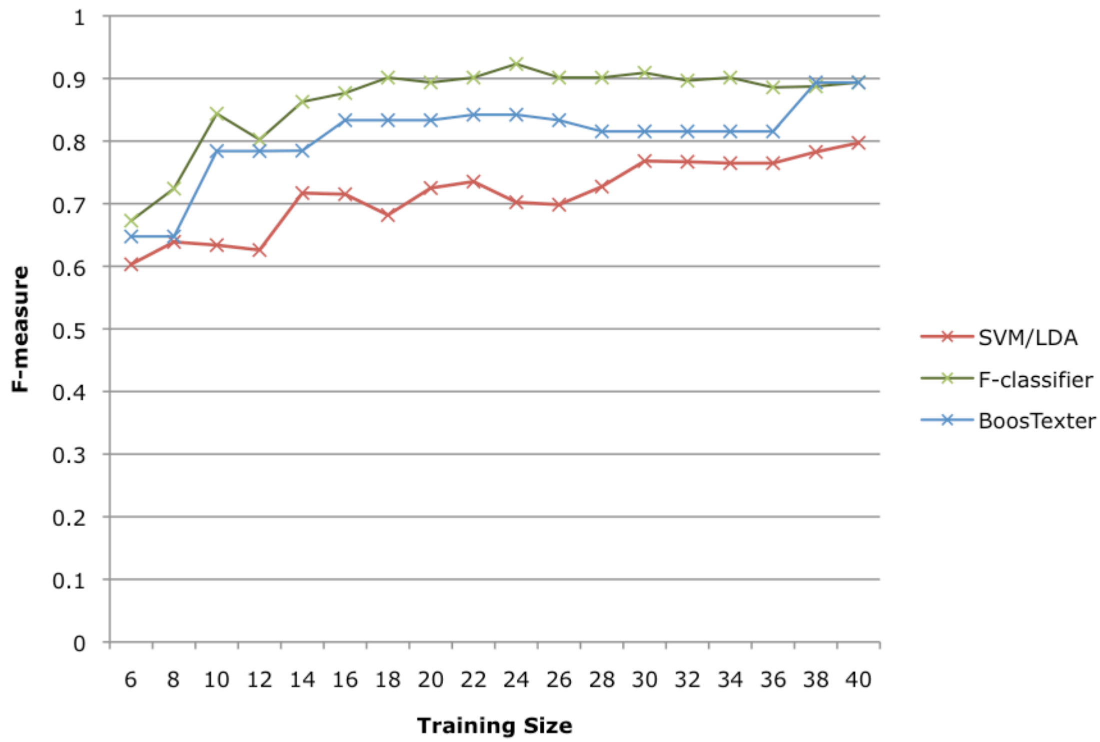


Figure 20. Obesity classifier performance

Figure 21 illustrates the F-classifier's performance as the number of clusters and the training set size was varied. For low training set sizes (n < 20), only the mid-ranged values for K performed well (10 <= K <= 24). At higher training set sizes (n >= 22) all values for K greater than 10 performed well.

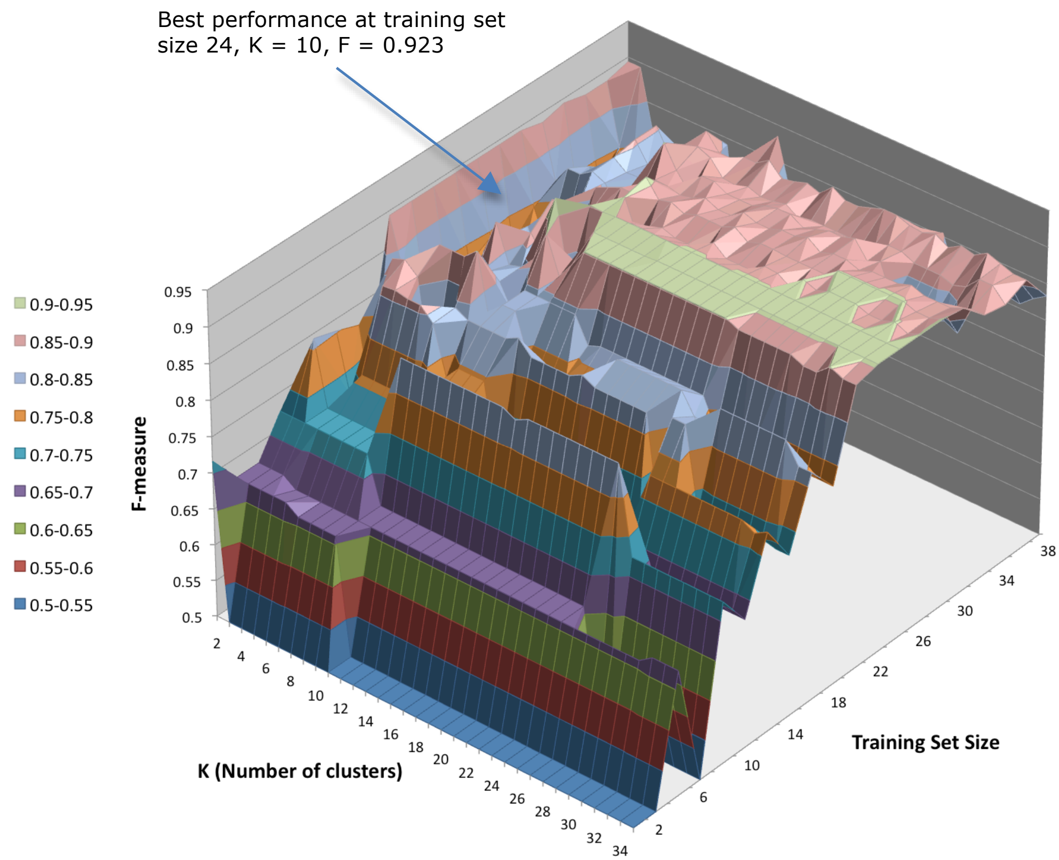


Figure 21. Obesity F-classifier performance contour map

<u>Trained feature performance</u>

Table 15 lists the AUC and Student's *t*-test results for the obesity F-classifier's trained features (features based on training data). Six of the ten features were good indicators of class for relevance to obesity based on these measures. This included both the nLM, and LDA features, as well the 4 NTPs *[weight]* (*p*-value<0.001;*AUC*=0.790), *[obes]* (*p*-value<0.001;*AUC*=0.778), and *[overweight]* (*p*-value<0.001;*AUC*=0.746). One additional NTP, *[glucos]* (*p*-value<0.001;*AUC*=0.556) performed well based on the *t*-test, but not as well on AUC. All NTPs with the exception of *[adolesc]* were conceptually related to the topic of obesity.

Feature	Mean score of GS articles: Positive	Negative	*p*-value	*AUC*
nLM	378.333	576.161	**< *0.001***	**0.729**
LDA	1.304	1.934	**< *0.001***	**0.914**

NTPs	Mean score of GS articles: Positive	Negative	*p*-value	*AUC*
[adolesc]	91.591	96.611	0.237	0.524
[bodi]:[mass]:[index]	71.972	100.000	**< *0.001***	**0.641**
[food]:[intak]	95.797	100.000	0.112	0.521
[glucos]	88.775	100.000	**< *0.008***	0.556
[obes]	40.951	94.933	**< *0.001***	**0.778**
[overweight]	50.889	100.000	**< *0.001***	**0.746**
[waist]	94.390	100.000	0.065	0.528
[weight]	33.998	91.550	**< *0.001***	**0.790**

Table 15. Obesity F-classifier feature performance

Classification examples

The F-classifier only produced 5 type II, and 6 type I errors. Type II errors were a result of news articles that discussed obesity but lacked any prominent terms or phrases directly talking about obesity. Examples included a news article discussing reduced exercise capacity[*], and a news article discussing how an improved diet can result in improved school performance[†].

The F-classifier's 6 type I errors were produced when terms typically associated to a discussion of obesity appeared repeatedly or prominently in the news article not primarily about obesity. Two examples of this were a news article discussing a link between nonfat milk and prostate cancer[‡], and another news article about the hazards of chronic sleep loss[§]. The low-fat news article repeatedly used terms such as *fat*, and *weight*, while in the

[*] Full text on page 220.
[†] Full text on page 221.
[‡] Full text on page 222.
[§] Full text on page 223.

chronic sleep loss news article, obesity is mentioned as a health problem associated to the primary news topic of chronic sleep loss.

BoosTexter, because of its reliance on rules derived from the appearance of words, made similar keyword based type I and type II errors. For example BoosTexter made type I errors on a news article discussing preterm babies being at risk for hospitalization as adults[*], and on the nonfat milk article that caused a type I error in the F-classifier. These errors were caused by the presence of phrases discussing nutritional problems and birth weight in news articles not about obesity. One example of a BoosTexter type II error was a news article discussing gastric bypass surgery[†]. BoosTexter likely classified this as not being about obesity because the news article did not possess any terms or phases that specifically discussed the surgery's role in treating obesity.

The keyword obesity classifier had the best recall of all classifiers. Most of its misclassifications were type I errors caused by the appearance of keywords in news article titles not predominantly about obesity. Examples of news articles that caused these errors included the news article discussing preterm babies that also produced a type I error in the BoosTexter classifier, and the news article about nonfat milk that resulted in type I errors for both the F-classifier and BoosTexter.

The SVM/LDA classifier had the lowest estimated performance. In addition, its errors were the most disparate of the three. There were a number of news articles that all other classifiers correctly categorized that were incorrectly classified by the SVM/LDA classifier, or vice versa. This was likely due to the SVM/LDA classifier's lack of dependence on the simple appearance of keywords. For example SVM/LDA classifier was the only classifier that correctly classified the nonfat milk news article as not about obesity. Another example of a non-obesity news article that only the SVM/LDA classifier correctly categorized, was one

---

[*] Full text on page 225.
[†] Full text on page 227

discussing a survey of the side effects of prostate treatments[*]. The word *obesity* appears in the first sentence of that news article, but the news article is not about obesity. The downside to this is that the SVM/LDA classifier also misclassified news articles that all other classifiers got correct. One example was a news article discussing how increases in throat cancer appear to be paralleling the obesity rate[†]. This news article repeatedly uses the word *obesity*, but the numerous other biological concepts likely influenced the SVM/LDA classifier towards its misclassification.

8.3. Lifestyle framing classification results

<u>Overall performance of lifestyle framing classifiers</u>

The F-classifier was the best performing classifier based on F-measure for the lifestyle framing task with an F-measure of 0.8 . Optimal performance was achieved with a training set of 10 news articles, an LDA index with 200 topics, and the *K*-means algorithm generating 9 clusters. The SVM/LDA classifier had the second best performance with an F-measure of 0.738 . This was achieved with an LDA index of 150 topics and a training set size of 38 news articles. BoosTexter had the lowest performance with an F-measure of 0.72 . This was accomplished with 1000 rounds of training using both sparse and standard trigrams, and a training set size of 22. The only statistically significant difference across all estimated classifier performances was that the F-classifier outperformed BoosTexter in recall by a margin of 0.225 (p-value=0.011).

[*] Full text on page 229.
[†] Full text on page 240.

Classifier	**Precision**	**Recall**	**F**
News Feature	0.720	0.900	0.800
BoosTexter	0.771	0.675	0.720
SVM/LDA	0.705	0.775	0.738

Classifiers	**Measure**	**Difference**
F-classifier vs. BoosTexter	Precision	-0.051 (*p*=0.329)
	Recall	**0.225 (*p*=0.011)**
	F-measure	0.080 (*p*=0.150)
F-classifier vs. SVM/LDA	Precision	0.015 (*p*=0.829)
	Recall	0.125 (p=0.179)
	F-measure	0.062 (p=0.291)
SVM/LDA vs. BoosTexter	Precision	-0.067 (p=0.422)
	Recall	0.100 (p=0.481)
	F-measure	0.018 (p=0.833)

Table 16. Lifestyle classifier results

Figure 22 illustrates the trends in performance of the 3 classifiers as the training set size increased from 6 to 40. Starting with a training set size of 6, the 3 classifiers have very different F-measure values. BoosTexter starts with the lowest performance, slightly above 0.55. Slightly better is the SVM/LDA classifier with an F-measure slightly above 0.6. The F-classifier estimated performance is slightly below 0.75 . From there the F-classifier's performance did not change appreciably, but was consistently higher than the other classifiers. The SVM/LDA classifier's performance showed small but constant increases in performance as the training set size increased. BoosTexter's performance was the most erratic of the three, rising and falling dramatically at the lower and higher training set sizes, and maintaining a mostly stable F-measure of 0.7 throughout the middle range of training set sizes tested (12 <= *n* <= 32).

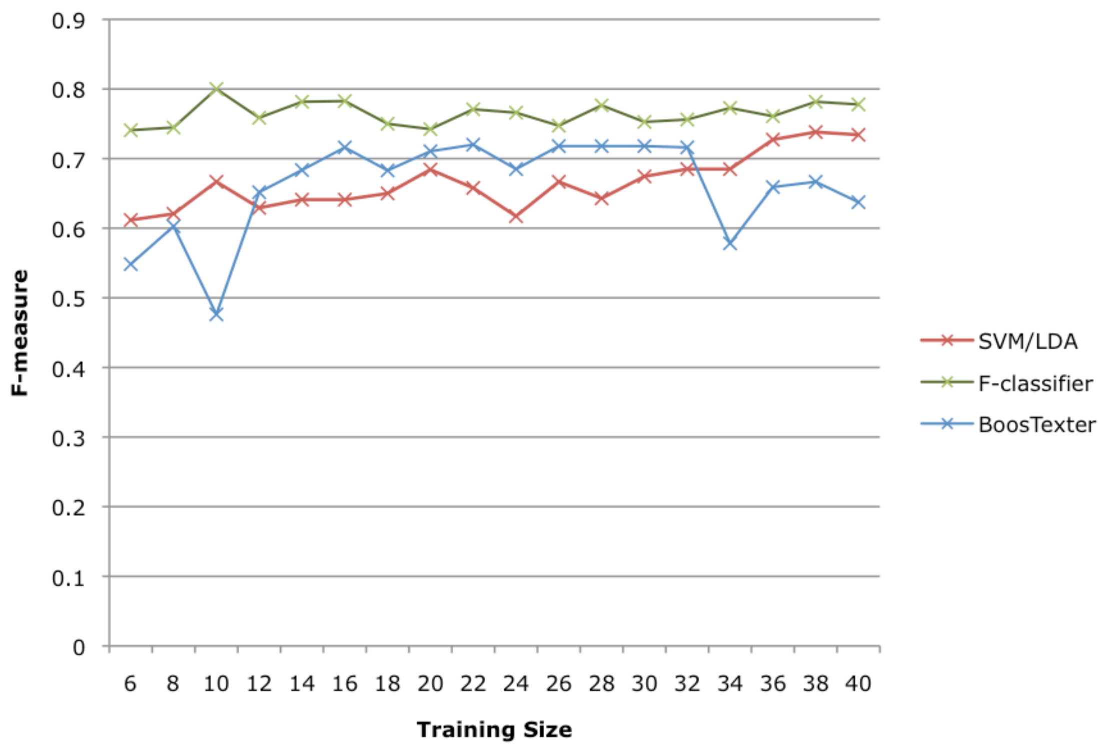


Figure 22. Lifestyle framing classifier performance

The contour map in Figure 23 illustrates how the F-classifier's performance varied for the lifestyle framing task as the training set size and the number of clusters varied. As opposed to the trend observed for the obesity F-classifier (see page 104) here there was no observable overall trend as the number of clusters was varied. Instead various clustering values produced performance near the optimal F-classifier's performance (F-measure between 0.75 and 0.8).

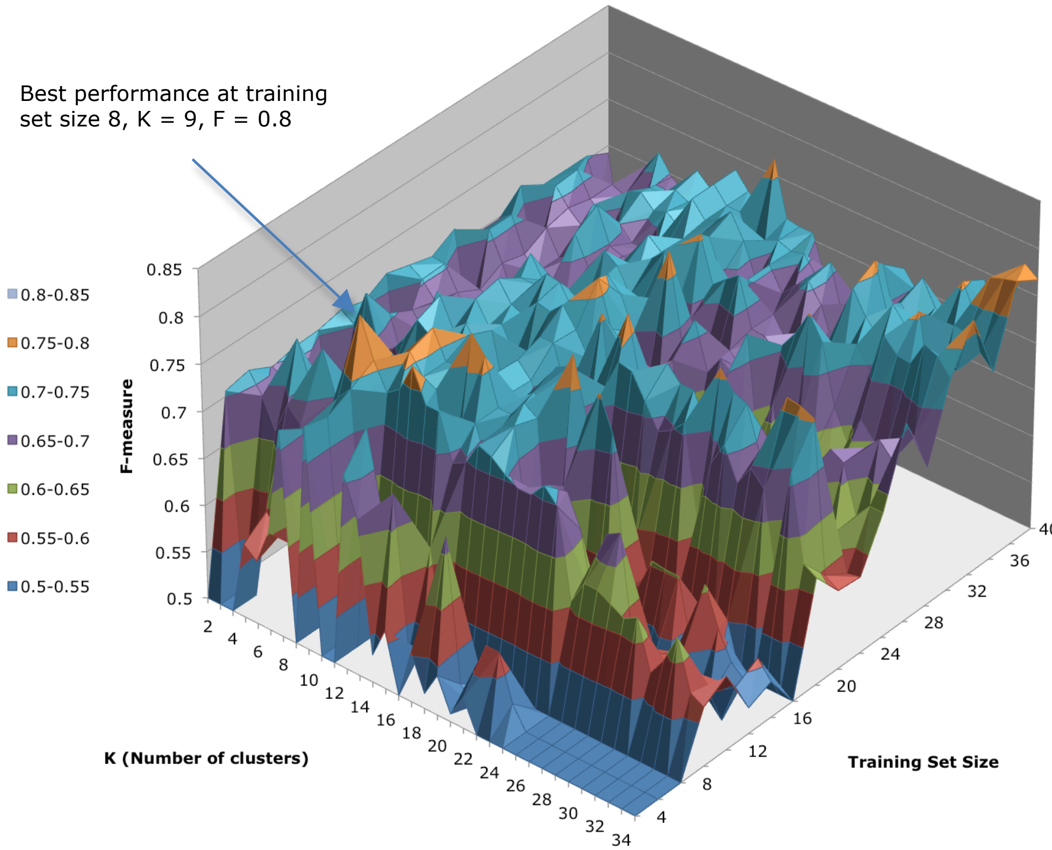


Figure 23. Lifestyle framing F-classifier performance contour map

<u>Trained feature performance</u>

For the lifestyle news framing task LDA was a good indicator of class (p-value=0.023; *AUC*=0.642), as were 3 of the 8 discovered NTPs; *[diet]* (p-value=0.043; *AUC*=0.66), *[exercise]* (p-value=0.009; *AUC*=0.663), and *[lifestyl]* (p-value=0.028; *AUC*=0.613). However, while these 4 features possessed good t-test, and AUC scores, none of them had *very* good scores (i.e. p-values < 0.001 or *AUC* > 0.7).

	Mean scores of GS articles			
Feature	**Positive**	**Negative**	***p*-value**	***AUC***
nLM	277.869	287.375	0.647	0.551
LDA	1.553	1.721	***0.023***	**0.642**

	Mean scores of GS articles			
Stemmed *n*-gram	**Positive**	**Negative**	**p-value**	***AUC***
[activ]	0.009	0.003	0.082	0.598
[care]	0.002	0.001	0.569	0.536
[chang]	0.003	0.003	0.801	0.514
[diet]	0.014	0.003	**0.043**	**0.660**
[exercis]	0.029	0.007	**0.009**	**0.663**
[lifestyl]	0.009	0.001	**0.028**	**0.613**
[obes]	0.016	0.018	0.744	0.527
[spend]	0.001	0.000	0.159	0.525

Table 17. Lifestyle framing news classifier results

Classification examples

Many news stories framed as lifestyle issues (in the gold standard) discussed diet or exercise. These news stories were the easiest to automatically classify successfully. Examples included a news article about how exercise and limited TV can decrease obesity in children[*], and another discussing studies discussing the impact of exercise in the obese and those diagnosed with diabetes[†].

The automated classifiers were also easily able to categorize news articles that focused exclusively on biology or medical interventions not related to lifestyle change as not framed as lifestyle stories. Examples included a news article discussing how a worsening in

[*] Full text on page 208
[†] Full text on page 243

incontinence has not been linked to menopause[*], and another describing how multivitamins are a popular dietary supplement among teenagers[†].

The classifiers did have difficulties when news articles combined terms and concepts related to lifestyle issues with ones related to other possible framings (e.g. biological causation). Such news articles produced unpredictable classification results, and often illustrated the strengths and weaknesses of the 3 classifiers. For example a news article titled "*Workouts boost function of insulin-making cells*"[‡] repeatedly discussed the benefits of exercise in the context of the beneficial biological processes that result from exercise. In this case BoosTexter was able to navigate the mixture of terms present in the news article to arrive at a correct classification of not framed as lifestyle, while the F-classifier and SVM/LDA classifiers misclassified this news article as a lifestyle story. The F-classifier and SVM/LDA classifiers' type I error was likely due to the news article text containing lifestyle related terms.

The F-classifier's recall was much better than the other two classifiers, and significantly better than the BoosTexter classifier (difference of 0.225; *p*-value=0.011). Over the entire lifestyle framing gold standard the F-classifier only produced 4 type II errors. The first was the news article discussing the parallel in the increase of throat cancer and obesity rates[§], the second was titled "*Americans losing sleep over financial crisis*[**], the third discussed a doctor group that supported trans-fat bans[††] and the last was titled "*Red yeast rice, fish oil fight high cholesterol*"[‡‡]. These examples were also problematic for the gold standard coders, one of whom commented that they

---

[*] Full text on page 231.
[†] Full text on page 218.
[‡] Full text on page 236.
[§] Full text on page 240.
[**] Full text on page 239.
[††] Full text on page 239.
[‡‡] Full text on page 237.

"*struggled with this one"*. Even after two rounds of coding, the coders did not unanimously agree on the coding of the trans-fat and financial crisis news articles.

8.4. Biological/genetic causation framing results

Overall performance of biological/genetic causation framing classifiers

Performance for biological/genetic causation framing classification was the lowest of the three classification tasks. The F-classifier's estimated F-measure performance of 0.755 was highest of the three classifiers. It achieved this performance using a training set of 42 news articles, an LDA index with 150 topics, and 12 clusters. The SVM/LDA classifier's best performance (*F*=0.709) was achieved with an LDA index of 75 topics and a training set size of 32. BoosTexter's best performance (*F*=0.634) was achieved using 10,000 and 100,000 iterations with standard trigrams and a training set size of 22. The F-classifier had significantly better recall than the other two classifiers. No other statistically significant differences in performance were observed.

Classifier	Precision	Recall	F
News feature	0.638	0.925	0.755
BoosTexter	0.619	0.650	0.634
SVM/LDA	0.718	0.700	0.709

Classifiers	Measure	Performance
News Feature vs. BoosTexter	Precision	0.019 (p=0.774)
	Recall	***0.275 (p=0.007)***
	F-measure	0.121 (p=0.067)
News Feature vs. SVM/LDA	Precision	-0.080 (p=0.172)
	Recall	***0.225 (p=0.022)***
	F-measure	0.046 (p=0.463)
SVM/LDA vs. BoosTexter	Precision	0.099 (p=0.195)
	Recall	0.050 (p=0.823)
	F-measure	0.075 (p=0.333)

Table 18. Biological/genetic causation framing classifier results

Figure 24 illustrates the trends in performance of the 3 classifiers as the training set size increased from 6 to 48. The F-classifier and the SVM/LDA classifier had very similar performance throughout. BoosTexter consistently exhibiting lower and much more erratic performance as the training set size was increased.

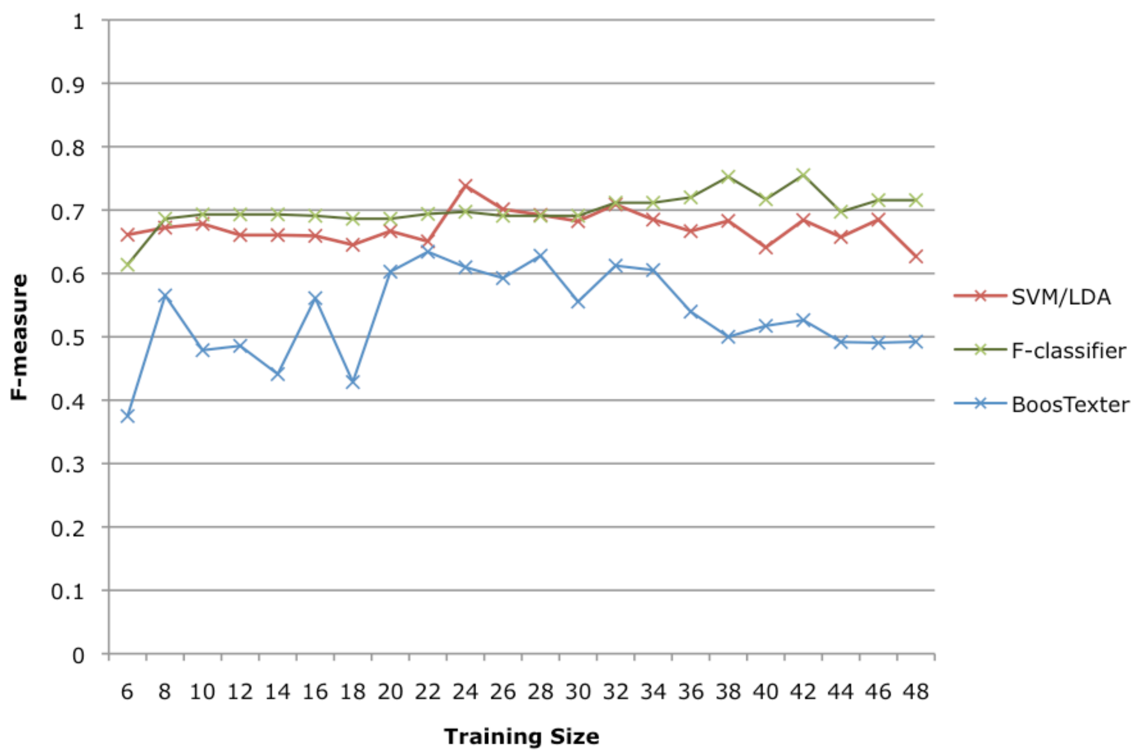


Figure 24. Biological/genetic causation framing classifier performance

Figure 25 illustrates how the genetic/biological framing F-classifier performance varied with changes in the training set size and the number of *K*-means clusters generated. From this chart we can see that the best performance was achieved with the combination of lower cluster sizes ($K<10$) and higher training set sizes ($n>30$).

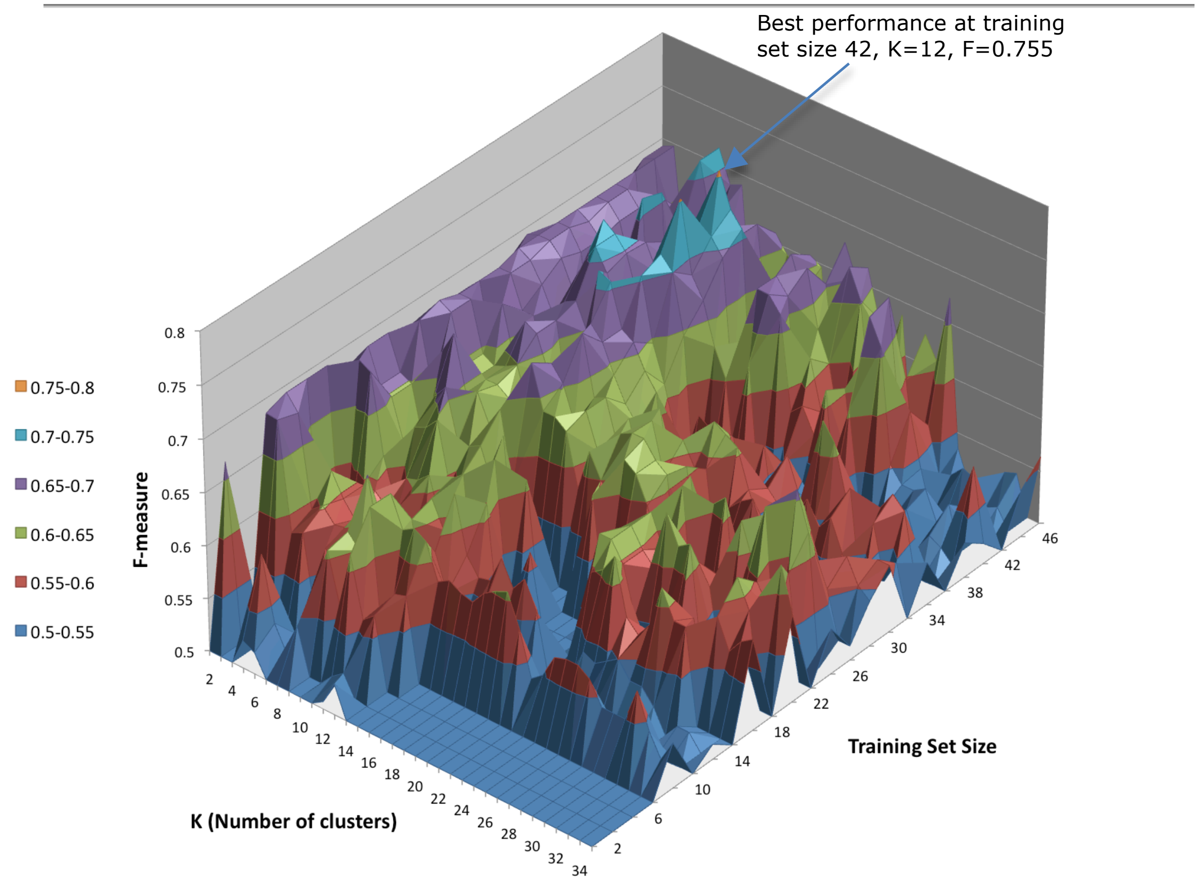


Figure 25. Biological/genetic framing F-classifier performance contour map

<u>Trained feature performance</u>

Only 3 of the possible 10 trained features proved to be good indicators of biological/genetic causation framing, and those 3 features performed only marginally well on the *t*-test and AUC measures (see Table 19). LDA and nLM had statistically significant differences between the mean scores of positive and negative cases, but neither had acceptable AUC scores, and while all NTPs were semantically related to genetic and biologic causation and/or obesity, only *[protein]* had a good AUC score (*AUC*=0.707).

	Mean score of GS articles			
Feature	**Positive**	**Negative**	***p*-value**	***AUC***
nLM	407.781	485.134	***0.028***	0.594
LDA	2.202	2.469	***0.006***	0.551

	Mean scores of GS articles			
NTPs	**Positive**	**Negative**	**p-value**	***AUC***
[cell]	0.005	0.000	0.100	0.563
[diet]	0.010	0.007	0.611	0.551
[exercis]	0.013	0.023	0.207	0.584
[fat]	0.012	0.009	0.602	0.579
[gene]	0.004	0.000	0.109	0.538
[genet]	0.003	0.000	0.083	0.550
[protein]	0.002	0.001	0.233	**0.707**
[weight]	0.021	0.015	0.309	0.550

Table 19. Biological/genetic framing F-classifier feature results

Classification examples

In general all classifiers were able to successfully categorize news articles that did not discuss any biological or genetic process as not biologically/genetically framed. For example all classifiers successfully categorized the news articles titled "*Lawmakers probe FDA approval of Ranbaxy drugs*"[*], and "*Internet helps doctor get back to basics*"[†] as not framed as biological/genetic causation health news. Another news article discussing obesity being seen as protective in some cases of heart failure[‡] was an example of a news story framed as biological/genetic causation that discussed numerous biological processes but lacked any strongly biological or genetic language. The F-classifier and the LDA/SVM classifier were able to correctly classify this as a biological/genetic causation health news story, while BoosTexter, with its sole reliance on the appearance of *n*-grams, produced an incorrect negative classification.

[*] Full text on page 232.
[†] Full text on page 233.
[‡] Full text on page 235.

The types of news articles that produced errors varied. The F-classifier had very high recall (0.925), with a much lower precision (0.638). This means that there was an increased number of type I errors and very few type II errors. The other classifiers had more even distributions of type I and type II errors.

One example of a news article that was unanimously judged by the coders as framed as biological/genetic causation, that produced a type II error for both BoosTexter and the SVM/LDA classifier was a news article discussing the link between worsening incontinence and menopause[*]. Examples of news articles that produced type I errors in at least 2 of the 3 classifiers included a news article about the observed parallels in throat cancer and obesity rate[†], which produced errors in BoosTexter and LDA/SVM classifiers, and the article discussing studies that show the benefits of exercise for obese individuals or those diagnosed with diabetes[‡], which resulted in a type I error for all classifiers.

8.5. Discussion

Results vs. Hypotheses

The results of this study support the hypotheses stated on page 80. The first hypothesis stated that there are patterns in the words and language of health news text that can be used to automatically classify health news by topic and framing. The estimated performance of the obesity news F-classifier *(prec=.917; recall=.93; F=.923)* strongly supports this hypothesis. The obesity F-classifier should be evaluated further to see if it generalizes to different news sources, different periods of time, and different health topics.

The F-classifiers developed for lifestyle framing *(prec=*.720*; recall=*.900*; F=*.800*)* and biological/genetic causation framing *(prec=*.638*; recall=*.925*; F=*.755*)* also performed well

[*] Full text on page 231
[†] Full text on page 240
[‡] Full text on page 243

enough to support the first hypothesis of this study. The framing classifiers demonstrated very good recall while still having an acceptable F-measure. However, their precision results must be improved before they can be put to practical situation.

The second hypothesis stated that classifying based on a news article's framing is a more difficult task than classifying based on broad health topics such as obesity. The lower agreement scores and coder comments generated during the development of the obesity and framing gold standards, along with the observed differences in classifier performance (framing classifiers not performing as well as obesity news classifiers) support this hypothesis.

The third hypothesis stated that acceptable performance could be achieved with very limited training data. The size of the training set required to achieve acceptable performance was an important part of this research because to make the use of classification algorithms practical for health news analyses, they must be trained with a minimum of training data. Figure 20, Figure 22, and Figure 24 on pages 103, 109, and 114 illustrate the impact that training set size had on the various classifiers. Overall, increases in training data did improve classifier performance, and for the small training set sizes of this study performance was good across all F-classifiers. Due to how different each classification task was, a general rule on how much training data is required could not be established. Future studies should investigate if the results observed here apply to other health topics and health news framings.

<u>Gold standards and news article content</u>

As just stated the process of creating the gold standards yielded insight into the difficulties of the classification tasks. The main challenge encountered by coders during the obesity gold standard development was in deciding how much of a discussion of obesity was required for a news article to be classified as an obesity news story. This challenge was articulated in the following comments submitted by the resolving coder:

"…you might connect a Mediterranean diet with less obesity, but since the article isn't talking about that, i'm saying not related, but i'd understand that this might be interpreted differently by others" [see page 217 for news article text]

"*though you should drink non-fat milk if you're obese, still not about obesity*" [see page 200 for news article text]

"…obesity probably plays a role in pelvic floor disorders, something a clinician might know and think about, but this is not mentioned at all in the article, so i'm saying no as I feel it's unfair to read beyond what is written based on outside knowledge" [see page 207 for news article text]

For the framing classification tasks the coders and classifiers had to resolve what it means to frame a news article as a lifestyle and/or biological/genetic causation issue. Even with the instructions and definitions provided to the coders there were still news articles that were especially challenging to code. For the lifestyle framing classification task the following comments from the second round of coding express the challenge coders faced:

News article title : **_Multivitamins are top diet supplement for teens_**
[full text on page 218]
Classifying for: **_Lifestyle_**

Coder	***Comment***	***Classified as***
Ca:	*"Upon closer reading, this article really is only talking about the prevalence of dietary supplement consumption..."*	Negative
Cb:	*"Discusses attributes that influence teen use of supplements, segmenting by culture. The focus on what influences choices leads me to add now include 'lifestyle'."*	Positive
Cc:	*"I don't think taking supplements is a lifestyle issue (?)"*	Negative

News article title : **_France reports leveling childhood obesity rates_**
[full text on page 204]
Classifying for: **_Lifestyle_**

Coder	***Comment***	***Classified as***
Ca:	*"I'd argue that this article is really about the success of policy changes/programs in France (and not so much about the changes in children's lifestyle)."*	Negative
Cb:	*"The article focusses on policy /enivornmental changes that influences lifestyle choices."*	Positive
Cc:		Positive

These examples typify the difficulty coders had in classifying the more ambiguous lifestyle framing cases.

For biological/genetic causation news framing, the final agreement score and the estimated classifier performances were the lowest of the 3 classification tasks. These results suggest that the task of classifying news articles on biological/genetic causation news framing was more difficult than either of the other two classification tasks.

One example of a news article that was already pointed out as problematic for coders (see page 100) is the news article discussing pelvic floor disorders among women. Below are the coder comments for this news article, and another that was also difficult to code.

News article title : ***Pelvic disorders common among women: study***
[full text on page 207]
Classifying for: ***Biological/Genetic causation***

Coder	***Comment***	***Classified as***
Ca:	*"Prevalence of pelvic floor disorders without delving into too much biological detail"*	Negative
Cb:	*"Assuming discussion of physical symptions, physiology, and biological issues are included in the category of "genetic/Bilogical", I'm going to keep my answer."*	Positive
Cc:	*"Being a woman seems like a genetic issue."*	Positive

News article title : ***Weight-loss surgery won't "cure" sleep apnea***
[full text on page 206]
Classifying for: ***Biological/Genetic causation***

Coder	***Comment***	***Classified as***
Ca:	*"Not enough to be biological; really more talking about how obesity and OSA are related but does not go into enough biological detail for me."*	Negative
Cb:	*"Discusses the implications of weight loss and OSA, providing perspective on implicatnios of weight loss surgery on OSA. *Assuming discussion of physical symptions, physiology, and biological issues are included in the category of "genetic/Bilogical", I'm going to keep my answer."*	Positive
Cc:		Negative

These comments reveal the challenges of coding news articles based on the framing. Future studies should address these problems by either providing more explicit, very narrowly articulated instructions on how to identify news articles that are framed in a particular way. This will help future health news investigations, and will improve any gold standard that is developed to evaluate health news classifiers.

Classifier design

The F-classifier implementation used basic K-means clustering to identify groups of positive and negative cases. An alternative semi-supervised clustering method such as constrained K-means clustering might improve performance. As described earlier, constrained clustering places added constraints on what items can be placed in the same cluster. The addition of such simple constraint rules derived from training cases might force the clustering algorithm to generate more accurate clusters when other features fail to do so.

Another technique to investigate is the use of LDA topic distributions as features. For the F-classifier, LDA was reduced to a single feature score. But LDA is a powerful and flexible method that can be levered in numerous ways. The SVM/LDA classifier was better than the other classifiers in certain cases. Future research should explore the use of LDA topics as a health news classification features, especially for the more conceptually challenging classification tasks.

F-classifier features

The second research question for this study asked what classification features would perform best. For classifying obesity health news all of the trained features proved to be good indicators of class. The LDA feature had the highest AUC score and a very good *t*-test result, as did the nLM features, and numerous NTPs scored well on both measures.

While few of the trained features scored well in the framing classifiers, there were many static features that were good discriminators of health news framing. For lifestyle framing all three DAL dimensions scored well on both AUC and *t*-test, as did article length. In addition 6 LIWC categories scored well on the AUC measure (see Table 12 on page 97). For biological/genetic causation framing 11 LIWC categories scored well (see Table 13 on page 99). In future studies these features should be given greater weight.

These results suggest that while classifying for overall health news framing is difficult, more targeted evaluations using specific health news framing features are possible. Also, in light of the comments made by the coders on the difficulty of the manual coding process, researchers should consider whether binary framing classifications of news article text should be avoided in favor of coding news articles on the specific attributes of a health news story that contribute to the framing of news coverage.

Selection of *K*

The obesity F-classifier was the least sensitive to changes in *K*. It demonstrated consistent performance for all values of *K* greater than 10 (see Figure 21). This means that valid groupings of items were maintained as the value of *K* was varied. One likely reason for this is that the difference in feature scores between positive and negative cases (obesity news and non-obesity news) was sufficient to overcome any sensitivity associated to changes in *K*.

The framing F-classifiers were more sensitive to changes in *K*, with different combinations of training size and *K* producing different levels of performance. However, the results charted in Figure 23 and Figure 25 suggest that the optimal value for *K* increased proportionate to the training set size, and the range of *K* that produced good performance was always less than the optimal *K*.

The range of values for *K* that produced good results for the framing F-classifiers can therefore be represented with the formula:

$$K <= CN$$

where *C* is a constant multiplier and *N* is the size of the training set. If this relationship is true it would mean that sensitivity to *K* for these classifiers may be associated with a lack of a enough positive training cases for each of the clusters of positive test cases.

These results need to be further analyzed and confirmed using more robust statistical methods, but from this data it is evident that the clustering results should be examined thoroughly, and future research should look to identify why different items are grouped together and how to improve overall classification performance.

Keyword filter, BoosTexter, and SVM/LDA classifiers

For the obesity news classification task the keyword filtering approach had the third best performance (based on F-measure) and the worst precision. Its recall was better than all other classifiers and was significantly better than all but the F-classifier. Based on these results researchers that rely exclusively on keyword filters to create a corpus of health news coverage for analysis should manually review all retrieved news articles and anticipate numerous type I errors.

BoosTexter, like the keyword filter, relies on the appearance of keywords in a text, but unlike keyword filtering BoosTexter automatically identifies rules based on training data. This enabled BoosTexter to outperform the keyword filters in precision in the obesity news classification task. But for the more complex framing classification tasks where keywords had a more limited impact on classifier performance, BoosTexter's performance lagged behind the other classifiers.

The SVM/LDA classifier provided substantially different results from the other classifiers. It had the lowest F-measure score for the obesity news classification task, but had good performance for the framing classification tasks, outperforming the other classifiers in most cases in precision, and in some cases in overall F-measure. These results suggest that the SVM/LDA classifier, and possibly other machine learning methods should be explored further for use in more conceptually difficult classification tasks such as news framing.

8.6. Summary

The goals of this study were to develop and evaluate text classification algorithms that could categorize health news based on relevance to obesity, and health news framing. Gold standards used to estimate classifier performance were created from the responses of expert coders. F-classifier (the classification software developed for this study) performance for the obesity news classifier was very good (supporting hypothesis 1 of this study). It outperformed the commonly used keyword filtering methodology most frequently used by health news researchers, and outperformed two other standard text classifiers (BoosTexter and SVM/LDA). The F-classifier should be evaluated further on different corpora, over other periods of time, and on other health topics.

F-classifiers that categorize based on health news framing were also developed and evaluated. Lifestyle and biological/genetic causation framing F-classifiers performed well overall (also supporting the first hypothesis of this study), but neither performed at a level that make them appropriate for practical use without human review of their output. These results support the second hypothesis of this study, which stated that the news framing classification task would be more difficult than classifying based on broader health topics. These results further suggest that the biological/genetic causation framing task is the more difficult of the framing tasks studied.

Some suggestions on how to improve the F-classifiers, in particular to improve performance on the news framing classification tasks, include switching to a constrained clustering method for grouping similar news articles, and exploring other classification features including expanded use of LDA topics as features.

While classifying based on health news framing proved to be very challenging, numerous static features of the new text such DAL and LIWC scores were good indicators of health news framing. These features should be studied further, and possibly used instead of overall binary framing classifications.

In the next chapter the design of the SalientHealthNews system will be described. SalientHealthNews combines algorithms, methods and techniques studied thus far in this dissertation with the findings of the literature review described in Chapter 4 to design a system that facilitates the use of computational methods to investigate health news coverage. To illustrate how SalientHealthNews can be used, preliminary data for a study of how health and obesity news coverage differs when various population are discussed will be presented.

Chapter 9. HEALTH NEWS RESEARCH SYSTEM: SalientHealthNews

This chapter describes the design of SalientHealthNews. SalientHealthNews is a system being developed to assist researchers in their investigation of health news coverage. The features of this system were defined based on the media effects research concepts described in Chapter 3, the findings of the literature review from Chapter 4, and the informatics and computational methods and concepts described in Chapters 5-8. In this chapter computational techniques that will be incorporated into the system are demonstrated, and other aspects of the system that are not yet implemented are described.

Along with the classification methods evaluated in the previous chapters the development and use of keyword filters that can categorize health news coverage by the populations discussed in news text is described. This type of filtering will immediately impact the usefulness of the SalientHealthNews system for investigators interested in examining how health news coverage differs based on what population is discussed.

This chapter will start by describing relevant concepts in the domains of data mining and hypothesis generation. This is followed by a description of the overall design and the major components of the SalientHealthNews system. Section 3 describes the use of SalientHealthNews computation methods in investigating how health and obesity news is framed when different populations are discussed. In the final section of the chapter I discuss how to incorporate other informatics, computation, and collaboration methods into SalientHealthNews to improve its usefulness and performance.

9.1. Data mining and hypothesis generation

The analysis of health news coverage is a data, and hypothesis rich domain. There is an avalanche of health news coverage published daily by traditional and new media heath news sources, and there are an endless number of media effects hypotheses that can be explored.

Unfortunately, all possible hypotheses cannot, and should not be fully investigated. Instead computational, data mining, information retrieval and statistical methods should be used to identify novel, relevant hypotheses that have an increased likelihood of producing significant findings. To do this health news researchers need to begin approaching the investigation of health news using data mining, multiple hypothesis testing, and hypothesis generation methods.

Multiple hypothesis testing and hypothesis generation

Hypothesis generation and the simultaneous testing of multiple hypotheses (MHT&G) are data driven approaches to the investigation and synthesis of novel and relevant ideas and hypotheses. They typically rely on a combination of data mining, information retrieval, visualization and the large-scale automated use of statistical hypothesis testing to discover statistically significant findings in vast amounts of data. These techniques did exist prior to the widespread use of computers, but as computers and computing power have become affordable and available to researchers in all domains of science, hypothesis generation and multiple hypothesis testing have become more widespread.

There are numerous scientific domains where these methods play an essential role in the selection of targets for study. Astronomy may be the oldest of these, where countless astronomical phenomena, and mathematical relationships were first observed in data, then hypothesized, then investigated further to produce supporting or contradicting data. In the biomedical sciences numerous examples of MHT&G exist. One prominent example is the use of gene expression microarray analysis to simultaneously perform thousand, or tens of thousands of gene expression studies[186]. Another example is the combining of microarray results with quantitative trait locus data to perform large-scale analysis of genotype-phenotype correlations[187]. Yet another example of a type that will likely gain in prominence as electronic health records become more common, is the data driven the discovery of adverse public health events such as pandemic influenza[188].

An example of MHT&G in bioinformatics that is more relevant to the research in this dissertation is text mining for literature-based discovery. Swanson is often credited for being one of the first researchers to successfully apply an exhaustive text mining methodology to biomedical literature in order to generate a testable hypothesis that later research confirmed[189]. In his seminal work Swanson sought to identify a treatment for Reynaud's disease, which at the time had no known cure. He performed an exhaustive search of research literature, first finding correlations in the literature between Reynaud's disease and vascular disorders. Then he identified correlations between vascular disorders that were similar to Reynaud's disease and fish oil. From this analysis of scientific literature, and not from any direct clinical research, he hypothesized that fish oil would either successfully treat, or cure Reynaud's disease[190]. This hypothesis was subsequently corroborated in clinical tests[191].

Since Swanson numerous biomedical scientific literature MHT&G methods and systems have been developed and are in use[192]. These systems use a wide range of natural language processing and data mining methods and constitute a very active field of research[193,194].

Finally, two examples of MHT&G tools that exist today, and leverage data and methods directly related to this dissertation, are Google Flu Trends[195], and HealthMap[196]. Google Flu Trends is a tool created by Google.org that uses aggregated Google search data to predict and estimate the intensity of influenza in different regions throughout the world. HealthMap uses the automated classification and visualization of a number of data sources, including WWW media reporting, to identify emerging global infectious disease outbreaks.

9.2. SalientHealthNews

The major components of the SalientHealthNews system are illustrated in Figure 26. These modules manage the acquisition, information retrieval, coding, scoring, and statistical analysis tasks performed during health news investigations. By directly linking these tasks to computational and statistical methods SalientHealthNews will facilitate current investigations of health news coverage, and provide researchers with the tools needed to approach health news analysis as a data mining and MHT&G task.

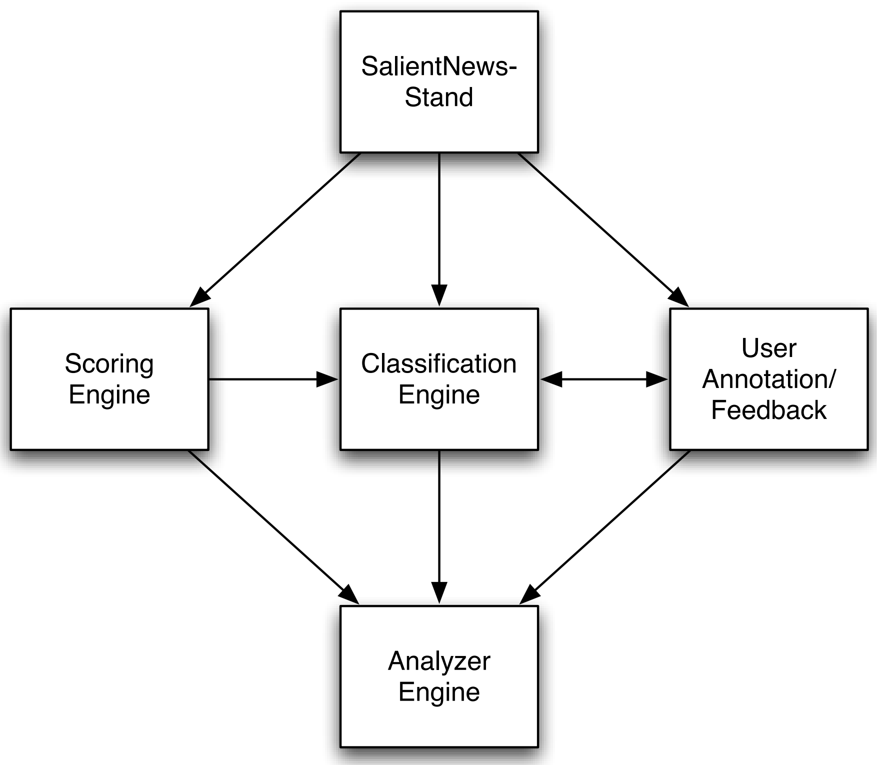


Figure 26. SalientHealthNews components

The first module is the *SalientNewsStand*. This module is based on the SalientNews RSS aggregator used to create the Reuter's health news corpus used in the study described in the previous chapter. Like the SalientNews aggregator, this module's function is to acquire, aggregate and prepare news text for downstream analysis. In addition to SalientNews' functionality SalientNewsStand will provide a user interface for navigating the news content in the SalientHealthNews system, and will be linked directly to other SalientHealthNews modules to provide them with news articles and accompanying metadata. The *Scoring Engine* uses a catalog of algorithms and preprocessing functions to score news articles. The

Scoring Engine will output news article scores and metadata to the *Classification Engine* and *Analyzer Engine*. The *Classification Engine* uses news article text, metadata, scores provided by the *Scoring Engine*, and manually entered feedback from human coders to classify news articles. Classifications produced by the *Classification Engine* are fed into the *Analysis Engine* for statistical analysis, and the *User Annotation/Feedback* module to gather feedback from people on the quality of the classifications. The *User Annotation/Feedback* module is a WWW user interface where news articles can be presented to people for annotation, coding, commenting and feedback. This module takes as input news articles and metadata from the *SalientNewsStand*, and classification data from the *Classification Engine*, and sends user provided data back to the *Classification Engine* in the form of training data for future classification tasks, and to the *Analyzer Engine* for statistical analysis.

Foundational work on SalientHealthNews modules has been demonstrated throughout this dissertation. The *SalientNewsStand* module is based on SalientNews[172], the RSS news aggregation system used to acquire the Reuter's health news corpus used in the preliminary study of Chapter 6, and in the classifier development and evaluation study of Chapters 7 and 8. The core of the *User Annotation/Feedback* module will be the software developed for the manual coding that occurred during the development of the obesity news and framing gold standards used in the classification study of the previous 2 chapters (refer to Appendix B on page 187 for screenshots of the WWW interface). The *Scoring Engine* and *Classification Engine* will utilize the algorithms developed and evaluated in the classification study (e.g. F-classifier). The *Analyzer Engine* will be based on the algorithms and software developed to facilitate the statistical analyses performed during the evaluation of the classification algorithms in the previous chapters.

9.3. Investigating population health news with SalientHealthNews

To demonstration SalientHealthNews' Scoring, Classification, and Analysis engines, an preliminary investigation of how framing features may vary based on what population is discussed in health news coverage was performed. The data generated by this investigation in presented in this section.

How do the media handle population specific health news?

Numerous studies have found observable differences in how the media cover health issues based on what population is being discussed, examples include Caulfield's investigation of representations of race in news coverage of BiDil[30], Beng's study on how gender is portrayed in news coverage of depression[109], Cohen's study comparing Cancer news coverage between black and general-audience newspapers[80], Gough's investigation of how men and women are portrayed in news coverage of dieting and nutrition, and Stryker's study of Cancer risk communications in mainstream and ethnic newspapers[197]. These studies tend to report mixed, conflicted findings. Some studies find that news coverage tends to propagate negative stereotypes and misconceptions regarding a population's attitudes, health status, or behaviors that impact health. But just as often (and in some instances simultaneously) studies find that news coverage caters to the specific health needs of a population, and discusses relevant health issues in positive, beneficial ways.

Much more research should be performed in this area. In particular it is essential that health news discussing populations with known health disparities be investigated to understand what the media is representing as important health issues for these populations, and how the media is communication health news when discussing these populations.

<u>Classifying news articles for obesity and population</u>

The SalientHealthNews Classification engine currently supports two classifiers. The first is an extended version of the basic keyword filter used by most health news researchers*, the second is the F-classifier developed in the classifier study of the previous 2 chapters. Both of these classifiers are abstract, requiring that numerous parameters be specified before the classifiers can be used to categorize news articles. The keyword filter requires that a keyword query be specified along with a minimum appearance threshold. The F-classifier requires that all its free parameters be specified. For this investigation the parameters used in the optimal obesity F-classifier (refer to page 101) were used.

Population news filters for nine populations were created using the extended keyword filter. Those populations were; women, men, children (including adolescents), seniors, African-American/black, Hispanic/Latino, gay/MSM/homosexual/bisexual (GMHB), people from rural communities, and people with the low socio-economic status (SES). In addition, the news articles for filters representing populations with significant documented health disparities (African-American/black, Hispanic, low SES, rural populations) were combined to form a separate, combined corpus of news coverage discussing populations with health disparities.

News articles were coded with a keyword filter's population if the filter matched 2 or more sentences from a news article, or one sentence and the news article's title. To assess the quality of the filters 100 news articles automatically labeled using the population filters were manually reviewed, and a number of randomly selected news articles from the remainder of the corpus were also reviewed.

* The SalientHealthNews keyword filter incorporates a number of searching features often not included in typical keyword searching tools; including support for regular expressions and searching for stemmed keywords.

As previously discussed keyword filters are not ideal, however they are acceptable in this investigation for two reasons; (1) keyword filters label text based on the actual appearance of words and terms, if a classification task can reasonably be reduced to the specific appearance of a select list of terms, as in this investigation, then the use of keywords is less problematic, (2) this investigation is being performed to produce *preliminary* data in the course of hypothesis generation. If more rigorous methods are needed they should be applied at a later stage.

The SalientHealthNews scoring module was used to score all Reuter's health news articles on all static dimensions identified in the previous study as being good indicators of class for the lifestyle and biological/genetic causation health news framings. Two LIWC categories not included in the original framing F-classifiers that are included here are the LIWC *home*, and *leisure* categories. These were included because the *leisure* category was subsequently identified as a good indicator of lifestyle framing, and *home* was a good indicator of both lifestyle and biological/causation framing. All static dimensions used in this investigation, along with the framing analysis in which they were used are listed in Table 20. Examples of words from these categories are provided in Appendix C on page 197.

	Features	**Lifestyle**	**Biological/genetic causation**
	Article length	X	
DAL features	*activation, imagery, pleasantness*	X	
LIWC features	*friend, certainty, home*	X	
	humans, leisure, ingestion	X	X
	biological processes, body, causation		X

Table 20. SalientHealthNews framing dimensions

The gold standard news articles identified for the classifier evaluation study described in the last two chapters was used as a baseline for average feature scores of news articles framed as lifestyle and biological/genetic causation. Student's *t*-test was used to identify significant differences between these baseline scores and the average feature scores of news articles discussing different populations. In addition to identifying average scores that differ significantly from the baseline established by the gold standard, SalientHealthNews was also used to identify if differences were in a negative or positive direction. A negative direction was a difference in feature score in the direction moving opposite the trend toward positive framing classifications, and a positive direction is a difference in the same direction as positive cases. For example if the average length of positively classified lifestyle framing news articles was 100, and the average of negatively classified news articles (those not framed as a lifestyle story) was 110, then the average news article length for a population filter that is positively significant would be less than 100 (because the trend indicates that lifestyle framed news articles are shorter than non-lifestyle news articles), while an average score greater than 100 would be negatively significant (i.e. in the direction of the negative cases). This should become clearer to the reader in the examples provided in the next section.

<u>Classification results</u>

The obesity F-classifier categorized 13%, or 432 out of the 3227 news articles in the Reuter's health news corpus as discussing obesity. The following table lists keyword filters, and the number of news articles that received positive classifications for each filter.

Population	Filters[*]	All health news n=3227	Obesity news[†] n=432;13.4%
African-American / Black	african*American or (black**(adult or women or men or children or patient)) or "blacks"	55 (1.7%)	9 (16.4%)
Children (+ adolescents)	*child* or babi or infant or boi or girl or newborn or kid or teen*	697 (21.6%)	125 (18.1%)
Elderly	elderli or (older (men or women or peopl or adult or patient)) or ((men or women or peopl or patient or adult or ag) [6-9][0-9][‡] (and or or) older)	101 (3.1%)	14 (13.9%)
GMHB	gai[§] or (homo or bi)sexu or men sex men[**] or msm	12 (0.4%)	0 (0.0%)
Hispanic / Latino	hispan or latin(o or a) or spanish	25 (0.8%)	6 (24%)
Men	man or men or boi or male or father*	312 (9.7%)	70 (22.8%)
Poor/low SES	(low*incom) or (low*socioeco*) or poverti or "the poor"	32 (1.0%)	3 (9.4%)
Rural pop.	rural*	10 (0.3%)	2 (0.5%)
Women	women or woman or girl or femal or mother*	620 (19.2%)	122 (19.8%)
Health disparity populations		104 (3.2%)	18 (4.2%)

***Error! Reference source not found.**. SalientHealthNews Population keyword filter results (continued.)*

Walkthrough of results for Hispanic/Latino health news

Table 21 shows the results that SalientHealthNews generated for the Hispanic/Latino filter. To help the reader understand how to interpret this data, I will walk through the different columns and rows, explaining each in turn. This will be followed by a summarization the data in this table.

---

[*] * = Wildcard match of 1 or more non-whitespace characters.
** = Wildcard match for 1 or more of any character (including whitespace)
Quoted words and phrases (e.g. "*blacks*") are literals, not word stems.

[†] The number of obesity news articles and the percentage of overall health news articles coded positively by each filter.

[‡] Capture all numbers *60 <= x <= 99.*

[§] Stem for the word "*gay*".

[**] Captures the phrase *men who have sex with men*.

The first column of Table 21 lists all the features included in the analysis. The first row of data is for the news article length feature (in characters). The next three rows are for the three DAL dimension features. The remaining rows are for the various LIWC category features. The first three columns of data combine to form a super-column containing analysis results over all health news articles that discussed the Hispanic/Latino populations. The second 3 columns combine to form another super-column, this time containing results from analyzing only obesity news coverage that discussed Hispanic/Latino populations. The 2 *mean* sub-columns are the average score over all news analyzed in that super-column for that row's dimension. The lifestyle framing column contains trend and *t*-test results comparing the average score for the news of that super-column (listed in the column to its left), to the average score of the positively classified lifestyle framing news articles from the framing gold standard form the classifier evaluation study of the previous chapters (which is used as a baseline). The biological/genetic causation framing columns are the same but for biological/genetic causation. *T*-test result cells are highlighted if the *p*-value indicates a statistically significant difference in average feature score (*p*-value<0.05). A green highlight indicates that the significant difference is in the same direction as going from the average of negative cases to positive cases for that framing. A highlight of pink indicates that the significant difference is in the direction of negative cases. No highlight indicates that the average is not `ly different than positively classified gold standard cases for that framing. Accompanying the *t*-test *p*-value is an indicator showing if the average score for that dimension, is greater than (⇑) or less than (⇓) the average of positive cases for that framing.

	News discussing Hispanics/Latinos Across all health news			News discussing Hispanics/Latinos Across obesity news		
Feature	mean	Lifestyle framing (trend) p	Bio/genetic framing (trend) p	mean	Lifestyle framing (trend) p	Bio/genetic framing (trend) p
length	1755.8	(⇓) 0.020		1876.7	(⇓) 0.322	
activation	1.336	(⇓) 0.079		1.388	(⇑) 0.468	
imagery	1.221	(⇓) 0.007		1.283	(⇑) 0.773	
pleasantness	1.425	(⇓) 0.118		1.518	(⇑) 0.043	
friend	0.0026	(⇓) 0.736		0.0055	(⇑) 0.509	
humans	0.0199	(⇓) 0.348	(⇑) 0.435	0.0315	(⇑) 0.088	(⇑) 0.067
cause	0.0128	(⇓) 0.035	(⇓) <0.001	0.0143	(⇓) 0.446	(⇓) 0.120
certain	0.0062	(⇑) 0.661		0.0112	(⇑) 0.057	
bio	0.0744		(⇓) <0.001	0.0742		(⇓) 0.051
body	0.0127		(⇓) 0.004	0.0084		(⇓) 0.021
health	0.0101		(⇓) 0.001	0.03		(⇑) 0.880
ingest	0.0062	(⇓) 0.004	(⇑) 0.466	0.0119	(⇓) 0.426	(⇑) 0.503
leisure	0.0023	(=) 1.000		0.0047	(⇑) 0.544	

Table 21. SalientHealthNews feature score statistical results.

To assist the reader in reading the results in Table 21 I will walk through the results for the *cause* dimension. For the *cause* dimension, the results in the *all health news* super-column indicates; (1) the average score for the *cause* dimension across all health news that discussed the Hispanic/Latino populations was 0.0128, (2) this average score was significantly lower (indicated by the down arrow) than the average *cause* score for all news articles classified as lifestyle framed in the framing gold standard, and significantly lower for the average *cause* score for all news articles classified as biological/genetic causation framed in the framing gold standard, 3) the difference in average *cause* scores between Hispanic/Latino health news and lifestyle framed news is in the direction of lifestyle framed news articles (indicated by the green highlight), 4) the difference in average *cause* scores between Hispanic/Latino health news and biological/causation framed news is in the direction of non-biological/genetic causation framed news (indicated by the red highlight).

Findings and generated hypotheses

The preliminary results in Table 21 suggest that general Hispanic/Latino health news is not like biological/genetic causation framed health news along 4 of the 6 tested features known to be good indicators of biological/genetic causation, but is very similar to lifestyle health news 7 of the 10 features indicative of lifestyle framing. Hispanic/Latino obesity news has even more features in common with lifestyle framed obesity news on lifestyle features (9 / 10), and is much more like biological/genetic causation obesity news (5 / 6 features).

These preliminary results suggests that average health news articles discussing Hispanics do not typically discuss biological process, and therefore it is less likely that these news articles are framed as biological/genetic causation health news. However, these news articles do use terms associated with health news article framed as lifestyle issues, suggesting that health news discussing Hispanics tends to be framed as lifestyle health news. Also suggested by these results is that obesity news that discusses Hispanics tends to be framed both as biological/genetic causation health news and as lifestyle health news.

The entirety of preliminary results (refer to Appendix E on page 246) suggests numerous additional hypotheses. Some hypotheses worth exploring based on this preliminary data are:

1. Obesity news coverage is typically framed as both a biological/genetic causation issue and a lifestyle issue when specific populations are discussed.
2. Overall health news coverage discussing populations with health disparities tends to score significantly lower on features linked to biological/genetic causation framing. This suggests that health news discussing these populations may not discuss the biological/genetic factors of health as frequently as other factors.
3. When children are discussed in overall health news very few features suggest that the health news coverage is framed as a lifestyle issue, and the features associated with biological causation present a mixed picture. This suggests that the typical

health news story about children is not about lifestyle change, and discusses some biological/genetic causation concepts (e.g. causative logical arguments and words associated with eating and ingestion), while not discussing others (biological processes, the body).

The next step with this investigation is to study a larger volume of news articles. Expanding the window of time over which the Reuters health news corpus is based, or adding news coverage from other health news sources can accomplish this. Additionally, smaller units of analysis should also be explored (e.g. paragraphs and sentences) to investigate if these features differ based proximity to the specific location where the different populations are mentioned.

9.4. Next steps for SalientHealthNews

Crowd sourcing

Internet facilitated crowd sourcing can help ease the burden of recruiting, training, using and retaining human coders for developing gold standard classifications. It is now commonplace for content-based websites to query users for simple feedback on the relevance and quality of the content they provide. Sites such as Reddit.com and Digg.com pioneered crowd sourcing as a way to identify the most popular web content submitted by users. Reddit allows users to click up and down arrows associated with user submitted links on the site. This results in the raising and lowing of that link's prominence. Digg is even simpler, just providing its users with a "Digg" button that raises a link's prominence. Many other sites have adopted similar means of getting user feedback to assess and rank content (e.g. Facebook).

Another online service that provides even more direct feedback for arbitrary coding tasks similar to those required by health news investigators is Amazon's mechanical turk[198]. This service links companies, researchers and developers with on demand human resources for a range of coding tasks. SalientHealthNews can use similar methods to solicit user feedback on the relevance of health news coverage to various topics and concepts in order to generate gold standards and training sets for classifier development and evaluation.

SalientHealthNews as a data mining tutor

One possible reason that investigators rely so heavily on manual content analysis methods and simple keywords may be that they have not had sufficient exposure to computational and data mining techniques to know how best to incorporate them into their research. SalientHealthNews may be able to bridge this gap by providing a clear step-by-step approach to applying computational and data mining techniques to the different phases of the health news analyses.

The first part of health news investigations is retrieving relevant health news coverage. To facilitate the transition to more robust information retrieval methods SalientHealthNews will support the simple keyword searches that researchers most frequently use. The searching and retrieval cycle can be augmented to include suggestions generated by SalientHealthNews on how to improve keyword queries to increase precision or recall depending on the investigator’s preference. Suggested improvements can be based on identifying synonyms for entered keywords, and stemmed versions of keywords that may perform better than basic wildcard searching. Two methods already used in this dissertation, LDA and LIWC word categories, can be used to implement this function.

Alternatively, investigators might prefer to identify examples of news articles that are relevant to their research. If they choose this option they can produce training sets from which either a keyword query can be automatically generated (using the NTP algorithm or LDA), or a health news classifier can be trained to perform the desired information retrieval task.

Alternatively, for classifier training, researchers can be presented with a predefined catalog of classifiers that can be immediately trained to retrieve relevant health news, or investigators can construct a classifier by choosing from a catalog of classifier features such as LIWC, DAL, nLM, and NTPs, and by selecting a feature vector classification methods (e.g. clustering, SVM, naïve Bayes).

Lastly SalientHealthNews will provide numerous methods for evaluating classifiers and classification features. Initially these methods will be the 3 techniques (the permutation method, *t*-test, AUC) used in the classifier implementation and evaluation study from the previous chapters.

Reuse of data across studies

Using SalientHealthNews as a repository for health news classifiers, and content analysis data should facilitate the reuse of data and methods across health news analysis studies. Many studies in the literature review in Chapter 4 were on similar health topics, and coded news articles on similar dimensions. However there was very little reuse of information retrieval parameters (e.g. keyword queries), datasets, and scoring/coding methods. Reuse across studies would save investigators time and resources, and would allow researchers to directly compare datasets and results.

Lastly, by promoting the sharing and reuse of methods, tools and datasets, SalientHealthNews will promote better transparency and reproducibility. As in many other scientific domains, once a health news study is complete, the datasets used for that study often are never used again. The corpora and content analysis results from these studies should be made available for future study and evaluation, and for use as training sets and gold standards for the development and evaluation of classifiers and scoring methods.

Smarter news feeds

The problem of acquiring relevant health news content is not limited to researchers seeking to evaluate health news coverage. Health news consumers may be interested in news coverage of health issues that affect them directly, public health professionals may need to monitor news coverage of an emerging health crisis, health communications professionals may need to follow a health story in real time so that they can craft messages for release to the public, professionals and corporate entities (companies, individuals, researchers, etc…) may want notification of news coverage about their research, products or area of interest. SalientHealthNews can leverage the classifiers and scoring methods used by researchers to generate better news feeds and monitoring tools. These feeds can then be made available to other researchers, the public, and other platforms and systems that require feeds of relevant health news content.

9.5. Summary

In this chapter I described the design and potential use of SalientHealthNews, a system that would provide health news investigators with data mining tools to facilitate their research. An overview of the major components and features of the system was presented. To demonstrate some of the classification, scoring and analysis techniques that will be available in the SalientHealthNews system I presented some preliminary data for an investigation exploring the lifestyle and biological/genetic causation framing features of health and obesity news coverage that discusses different populations. Last was a discussion of the next steps in the design and implementation of the system.

In the final chapter of this dissertation I will revisit the aims and hypothesis of the research, discuss how the literature review, classifier study, and SalientHealthNews system addressed those aims, and discuss the scientific contributions of this dissertation.

Chapter 10. DISCUSSION & CONCLUSION

In this, the final chapter, I will briefly discuss some of the major findings and contributions of this dissertation research along with suggestions of future work that should be explored.

10.1. Health news analysis literature review

Aim and major findings

The aim of the health news analysis literature review done for this dissertation was to identify the methods typically used, and the health topics most frequently studied by health news investigators. The results of the review revealed that investigators rely heavily on simple keywords for retrieving relevant health news (80% of studies), and on manual content analysis methods for scoring and coding health news coverage prior to analysis (93% of studies). Very few studies used any alternative methods. This supported the first hypothesis of this study, which stated that keywords and content analysis methods were more frequently used than other methods.

Another finding of the literature review was that framing studies were the most common (63% of studies). This contradicts the second hypothesis proposed for this study, which assumed that agenda setting would be more common. The most likely explanation for this finding is that the definition of a news framing study is very broad, encompassing much of what researchers are likely to investigate when studying health news. This is in contrast to the definitions of the other two media analysis types, agenda setting and informing studies, which are narrower in focus.

The literature review also revealed that 39% of studies found that health news coverage was of poor quality, while only 9% of studies found that health news coverage was of good quality. This supports the third hypothesis of this review, which stated that researchers continue to find that the quality of health news is poor.

There are numerous possible explanations for why researchers continue to find that health news coverage is poor in quality. Journalists have stated that covering medical and scientific news is very difficult. Health news is particularly challenging because journalists have to navigate the need to inform the public about health issues with the practical needs of the news media and profit driven news sources[67,119,120].

In addition to these findings there were numerous other interesting results including; (1) there was a greater focus on studying media coverage of cancer, and illness in general, over other health issues, (2) there was a greater frequency of studies of health news about women or children over all other populations, and (3) the most frequently used method for analyzing data was simple descriptive frequencies (54% of studies).

Contributions and next steps

This review is a major contribution to the study of health news coverage because it provides a snapshot of (1) the methods researchers use to evaluate health news, (2) the research questions and health topics that are of interest to investigators of health news coverage, and (3) the findings of these studies. No similar literature review of this scope was found in the research literature.

The findings regarding investigators' reliance on keyword searching and manual content analysis methods is important because it illustrates the gap between what is available in the computation text mining, NLP, and IR domains, and what is being used to investigate health news coverage.

This review should be expanded in a number of ways. Additional reviewers should be recruited so that the review becomes more systematic and fit for publication as peer reviewed research. To ensure that the review is truly comprehensive, biomedical literature search engines other than PubMed should be used. And additional variables should be coded for, including the type of coding or scoring done in each study (binary, categories, on a scale), and the unit of analysis (e.g. news article titles, or full text).

10.2. Health news classification

Aim and major findings

The aim of the classification study was to develop and evaluate an obesity news classifier and health news framing classifiers. The two hypotheses for this study were (1) that there are patterns in the words and language of health news text that can be used to automatically classify health news based on topic and framing, and (2) that the framing classification task would prove to be more difficult than classifying based on broad health topics such as obesity.

The F-classifiers developed for this study performed well, despite the use of very limited training data. This supports the first study hypothesis. The obesity news F-classifier outperformed all alternative classification methods evaluated in the study. The biological/genetic causation and lifestyle framing classifiers also performed well and outperformed alternative methods, but did not perform well enough to make them ready for immediate use in health news investigations. This later result supports the second hypothesis of the study, that framing is the more difficult classification task. The comments provided by the coders recruited to develop the classifier evaluation gold standards also expressed the difficulties of creating binary classifications on health news framing, lending even more support to the second study hypothesis.

Although the framing F-classifiers should not yet be used for "real" investigations, numerous features used in their development, including the DAL dimensions, and LIWC categories, proved to be good indicators of a news article's health news framing, and should be used when investigating the framing of health news.

Contributions and next steps

These results are important because they reveal that some of the coding and classification tasks performed by researchers can be facilitated by computational methods.

The obesity classifier performed well enough to justify its use in classifying health news in real studies. This suggests that other classifiers that categorize health news coverage based on health topic, that use limited training material, can also be developed and used.

For the classification of health news based on framing the study results suggest that framing is sufficiently more complex that simple binary classifications of framing are difficult and possibly not appropriate for the methods explored in this dissertation. However, specific features (DAL dimensions, LIWC categories, NTPs, LDA) performed well as indicators of health news framing, suggesting that when studying health news using computational methods it will be more appropriate to use these narrowly defined features as opposed to developing binary classifiers.

The results of this study point to numerous research opportunities. The obesity F-classifier should be evaluated on other health news corpora, published by other news organizations, over different periods of time. It should also be trained and evaluated on other health topics including diseases such as cancer and diabetes, risk factors such as smoking, and other health topics such as the US healthcare reform. Further research on classifying health news based on framing should also be done. Additional features should be tested, features evaluated in this study that did not perform well should be discarded, and alternative algorithms for generating classifications based on features should be explored.

Lastly, computational methods that can be used to evaluate health news coverage based on the information included in the news text should be investigated. To explore the informing features of news articles, information extraction (IE) methods should be developed and evaluated. These methods must identify the types of information that appear in a corpus of health news coverage. Numerous NLP and statistical text mining methods exist that may be able to perform this. One method worth exploring is Barzilay's content modeling method[199]. This method models documents in a corpus using hidden Markov models. Barzilay's method and numerous others should be developed and evaluated for use in informing studies of the media's coverage of health.

10.3. SalientHealthNews and population health news framing

Discussion

SalientHealthNews builds upon the findings of the literature review and classification studies to define the features of a system that can assist health news investigators in the use of text data mining techniques to explore and generate hypothesis for studying media coverage of health. Once completed SalientHealthNews will promote the sharing and reuse of methods and datasets across studies, streamline the use of text data mining methods in the evaluation of health news, enable multiple hypothesis testing and hypothesis generation for the investigation of health news coverage, employ crowd sourcing techniques to establish gold standards, and facilitate the development of health news feeds and notification services that can be used by individuals, organizations and systems that need to be updated on health news coverage in real-time.

To demonstrate SalientHealthNews' modules and algorithms it was used to investigate the lifestyle and biological/genetic framing of health and obesity news coverage that discussed different populations. Instead of using binary classifications of framings, features found in the classification study to be good indicators of these framings were studied as dependant variables. Even with a limited dataset preliminary results suggest that obesity news is more frequently framed as lifestyle and biological/genetic causation than general health news, and that when populations with health disparities are discussed there tends to be less use of terms associated with the biological/genetic causation features.

Contributions and next steps

The development of SalientHealthNews should be completed and investigators should be invited to use it in their research. The functions and methods that will be supported by SalientHealthNews are needed in this domain because the methods traditionally used in this field cannot keep pace with the growth in the volume of health news published, and the changes in the sources and of health news and how health news is distributed. As researchers become more interested in other forms of online health news content the challenges of relying solely on methods such as manual content analysis and keyword searching will become more pronounced. SalientHealthNews will provide researchers with the tools they need to continue to investigate health news.

The investigation of the framing features of population health news will continue, and the findings suggested by the preliminary results will be explored further. News articles discussing the different populations should be compared not only to the framing gold standards, but also to each other to see the direct differences in framing features between news coverage of different populations. Smaller units of analysis should also be explored, for example features at the paragraph and sentence level should be investigated to find more granular differences. Smaller units of analysis could answer questions about how word use differs based proximity to the mention of different populations of health issues such as obesity. Lastly, this analysis should be performed over a broader health news corpus that includes other news sources.

10.4. Final words

This dissertation contributes to the study of the media's coverage of health substantively and methodologically. It provides a detailed description of how health news is currently studied, and presents computational methods and tools that can assist investigators in their research. Researchers with an interest in assessing media coverage of health topics face great challenges regardless of the health issues they are studying or the ways that they are measuring news coverage. As the volume of health news continues to grow, and the forms that health news coverage takes continues to evolve, researchers must look to new methods and techniques. Just as information technology has transformed the way that news is generated, published and consumed, it must also transform the way that it is evaluated. Informatics, data mining, and computational methods such as those investigated in this dissertation should be explored and integrated into health news investigators' research via tools and systems such as SalientHealthNews. By doing so researchers will be able to accurately, consistently and systematically study and evaluate the media's coverage of health, and communicate their findings in a timely fashion to health care professionals, health communications professionals, health news journalists and the public. This will result in more informed health journalists, healthcare providers, and healthcare consumers, ultimately improving individual and public health.

"The increasing number of publishers who are elevating their standards for health news and health advertising constitutes an encouraging sign of facilitating health progress" - Unknown author; 1926[1]

BIBLIOGRAPHY

1. Health News and Ethics. California and Western Medicine. 1926;24:365.

2. Steele WR, Mebane F, Viswanath K, Solomon J. News media coverage of a women's health controversy: how newspapers and TV outlets covered a recent debate over screening mammography. Women & health. 2005;41:83-97.

3. Nicholson R, Kreuter M, Lapka C, Wellborn R, Clark E, Sanders-Thompson V, Jacobsen H, Casey C. Unintended Effects of Emphasizing Disparities in Cancer Communication to African-Americans. Cancer Epidemiology Biomarkers & Prevention. 2008;17:2946-2953.

4. World Health Organization. Preventing suicide. A resource for media professionals. 2000. URL: http://www.who.int/mental_health/media/en/426.pdf (Accessed February 1, 2011)

5. Robinson M, Appel K. Network news coverage of Congress. Political Science Quarterly. 1979;94:407-418.

6. Antilla L. Climate of scepticism: US newspaper coverage of the science of climate change. Global Environmental Change. 2005;15:338-352.

7. Dearing J. Newspaper coverage of maverick science: Creating controversy through balancing. Public Understanding of Science. 1995;4:341.

8. Au JS, Yip PS, Chan CL, Law YW. Newspaper reporting of suicide cases in Hong Kong. Crisis. 2004;25:161-8.

9. Angermeyer MC, Schulze B. Reinforcing stereotypes: How the focus on forensic cases in news reporting may influence public …. International Journal of Law and Psychiatry. 2001;24:469-486.

10. Wilson A, Bonevski B, Jones A, Henry D. Media reporting of health interventions: signs of improvement, but major problems persist. PLoS ONE. 2009;4:e4831.

11. Hughes E, Kitzinger J, Murdock G. The Media and Risk. In: Risk in Social Science. Oxford: Oxford University Press; 2006:250-270.

12. Dorfman L. Studying the news on public health: how content analysis supports media advocacy. American Journal of Health Behavior. 2003;27 Suppl 3:S217-26.

13. Brown J, Chapman S, Lupton D. Infinitesimal risk as public health crisis: news media coverage of a doctor-patient HIV contact tracing investigation. Social science & medicine (1982). 1996;43:1685-95.

14. Moynihan R, Bero L, Ross-Degnan D, Henry D, Lee K, Watkins J, Mah C, Soumerai SB. Coverage by the news media of the benefits and risks of medications. New England Journal of Medicine. 2000;342:1645-50.

15. Rubin DM, Hendy V. Swine influenza and the news media. Annals of Internal Medicine. 1977;87:769-74.

16. Conrad P. Genetic optimism: framing genes and mental illness in the news. Culture, Medicine and Psychiatry. 2001;25:225-47.

17. Wenger L, Malone R, Bero L. The cigar revival and the popular press: a content analysis, 1987-1997. American Journal of Public Health. 2001;91:288-91.

18. Dalkey N. The Delphi method: An experimental study of group opinion. 1969, Rand Corporation

19. Kovic I, Lulic I, Brumini G. Examining the medical blogosphere: an online survey of medical bloggers. Journal of Medical Internet Research. 2008;10:e28.

20. Heilman JM, Kemmann E, Bonert M, Chatterjee A, Ragar B, Beards GM, Iberri DJ, Harvey M, Thomas B, Stomp W, Martone MF, Lodge DJ, Vondracek A, de Wolff JF, Liber C, Grover SC, Vickers TJ, Mesko B, Laurent MR. Wikipedia: a key tool for global public health promotion. Journal of Medical Internet Research. 2011;13:e14.

21. Barness LA, Opitz JM, Gilbert-Barness E. Obesity: genetic, molecular, and environmental aspects. Am J Med Genet A. 2007;143A:3016-34.

22. Health Implications of Obesity. NIH Consens Statement Online 1985 Feb 11-13. 1985, NIH

23. Lawrence R. Framing Obesity: The Evolution of News Discourse on a Public Health Issue. International Journal of Press/Politics. 2004;9:56-75.

24. Kim S-H, Willis LA. Talking about obesity: news framing of who is responsible for causing and fixing the problem. Journal of health communication. 2007;12:359-76.

25. Henderson J, Coveney J, Ward P, Taylor A. Governing childhood obesity: framing regulation of fast food advertising in the Australian print media. Social science & medicine (1982). 2009;69:1402-8.

26. Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care. Washington, DC, USA: National Academies Press; 2003

27. Michael Halle CBL, Meena Seshamani. Health Disparities: A Case for Closing the Gap. Journal. 2009;

28. Amzel A, Ghosh C. National newspaper coverage of minority health disparities. Journal of National Medical Association. 2007;99:1120-5.

29. Taylor-Clark KA, Mebane FE, Steelfisher GK, Blendon RJ. News of disparity: content analysis of news coverage of African American healthcare inequalities in the USA, 1994-2004. Social science & medicine (1982). 2007;65:405-17.

30. Caulfield T, Harry S. Popular representations of race: the news coverage of BiDil. The Journal of Law, Medicine and Ethics. 2008;36:485-90.

31. Kim AE, Kumanyika S, Shive D, Igweatu U, Kim S-H. Coverage and framing of racial and ethnic health disparities in US newspapers, 1996-2005. American Journal of Public Health. 2010;100 Suppl 1:S224-31.

32. Collins PA, Abelson J, Pyman H, Lavis JN. Are we expecting too much from print media? An analysis of newspaper coverage of the 2002 Canadian healthcare reform debate. Social science & medicine (1982). 2006;63:89-102.

33. Iyengar S. Overview. In: Do the media govern? Politicians, voters, and reporters in America. Thousand Oaks, CA: Sage Publications; 1997:211-216.

34. Iyengar S, Peters M, Kinder D. Experimental demonstrations of the "not-so-minimal" consequences of television news programs. American Political Science Review. 1982;

35. Theodoridis S, Koutroumbas K. Introduction. In: Pattern recognition. San Diego, CA: Elsevier; 2006:837.

36. Andersson CM, Bjärås G, Tillgren P, Ostenson C-G. Local media monitoring in process evaluation. Experiences from the Stockholm Diabetes Prevention Programme. Journal of health communication. 2007;12:269-83.

37. Barnes MD, Hanson CL, Novilla LMB, Meacham AT, McIntyre E, Erickson BC. Analysis of media agenda setting during and after Hurricane Katrina: implications for emergency preparedness, disaster response, and disaster policy. American Journal of Public Health. 2008;98:604-10.

38. Hilbert A, Ried J. Obesity in Print: An Analysis of Daily Newspapers. Obesity Facts. 2009;2:46-51.

39. Krippendorff K. Content analysis: An introduction to its methodology. CA: Thousand Oaks; 2004

40. Chau J, Bonfiglioli C, Chey T, Bauman A. The Cinderella of public health news: physical activity coverage in Australian newspapers, 1986-2006. Australian and New Zealand Journal of Public Health. 2009;33:189-192.

41. Halpin M, Phillips M, Oliffe JL. Prostate cancer stories in the Canadian print media: representations of illness, disease and masculinities. Sociology of Health & Illness. 2009;31:155-169.

42. Rock M. Diabetes portrayals in North American print media: a qualitative and quantitative analysis. American Journal of Public Health. 2005;95:1832-8.

43. Champion D, Chapman S. Framing pub smoking bans: an analysis of Australian print news media coverage, March 1996-March 2003. Journal of epidemiology and community health. 2005;59:679-84.

44. Calloway C, Jorgensen CM, Saraiya M, Tsui J. A content analysis of news coverage of the HPV vaccine by U.S. newspapers, January 2002-June 2005. Journal of women's health (2002). 2006;15(7):803-9.

45. Listerman T. Framing of science issues in opinion-leading news: international comparison of biotechnology issue coverage. Public Understanding of Science. 2010;19:5.

46. Stebbing C, Kaushal R, Bates DW. Pediatric medication safety and the media: what does the public see? Pediatrics. 2006;117:1907-14.

47. Smith EA. 'It's interesting how few people die from smoking': tobacco industry efforts to minimize risk and discredit health promotion. European Journal of Public Health. 2007;17:162-70.

48. Anderson AL, Leffler K. Bedbug infestations in the news: a picture of an emerging public health problem in the United States. Journal of Environmental Health. 2008;70:24-7, 52-3.

49. Haddock CK, Parker LC, Taylor JE, Poston WS, Lando H, Talcott GW. An analysis of messages about tobacco in military installation newspapers. American Journal of Public Health. 2005;95:1458-63.

50. Genovesi AL, Donaldson AE, Morrison BL, Olson LM. Different perspectives: a comparison of newspaper articles to medical examiner data in the reporting of violent deaths. Accident, Analysis and Prevention. 2010;42:445-51.

51. Hilton S, Hunt K. UK newspapers' representations of the 2009-10 outbreak of swine flu: one health scare not over-hyped by the media? Journal of epidemiology and community health. 2010;

52. Nelson DE, Evans WD, Pederson LL, Babb S, London J, McKenna J. A national surveillance system for tracking tobacco news stories. American Journal of preventive medicine. 2007;32:79-85.

53. Nichols S, Chase N. A content analysis of health research reported by the daily newspapers of Trinidad and Tobago. The West Indian medical journal. 2005;54:308-14.

54. Lewison G, Tootell S, Roe P, Sullivan R. How do the media report cancer research? A study of the UK's BBC website. British Journal of Cancer. 2008;99:569-76.

55. Atkin CK, Smith SW, McFeters C, Ferguson V. A comprehensive analysis of breast cancer news coverage in leading media outlets focusing on environmental risks and prevention. Journal of health communication. 2008;13:3-19.

56. Boak MB, M'ikanatha NM, Day RS, Harrison LH. Internet death notices as a novel source of mortality surveillance data. American Journal of Epidemiology. 2008;167:532-9.

57. Fishman J, Ten Have T, Casarett D. Cancer and the media: how does the news report on treatment and outcomes? Archives of internal medicine. 2010;170:515-8.

58. Moriarty CM, Jensen JD, Stryker JE. Frequently cited sources in cancer news coverage: a content analysis examining the relationship between cancer news content and source citation. Cancer Causes Control. 2010;21:41-9.

59. Schwitzer G. How do US journalists cover treatments, tests, products, and procedures? An evaluation of 500 stories. PLoS Medicine. 2008;5:e95.

60. Pirkis J, Dare A, Blood RW, Rankin B, Williamson M, Burgess P, Jolley D. Changes in media reporting of suicide in Australia between 2000/01 and 2006/07. Crisis. 2009;30:25-33.

61. Entman RM. Framing: Toward clarification of a fractured paradigm. Journal of Communication. 1993;43:51-58.

62. Scheufele DA. Framing as a theory of media effects. Journal of Communication. 1999;103.

63. D'Angelo P. News framing as a multiparadigmatic research program: A response to Entman. Journal of Communication. 2002;52:870-888.

64. Gamson WA, Modigliani A. Media Discourse and Public Opinion on Nuclear Power: A Constructionist Approach. The American Journal of Sociology. 1989;95:1-37.

65. Corrigan PW, Watson AC, Gracia G, Slopen N, Rasinski K, Hall LL. Newspaper stories as measures of structural stigma. Psychiatric services. 2005;56:551-6.

66. Kitzinger J. Media templates: patterns of association and the (re) construction of meaning over time. Media, Culture & Society. 2000;22:61-84.

67. Hodgetts D, Chamberlain K, Scammell M, Karapu R, Waimarie Nikora L. Constructing health news: possibilities for a civic-oriented journalism. Health (London, England : 1997). 2008;12:43-66.

68. Clarke J. Heart disease and gender in mass print media. Maturitas. 2010;65:215-8.

69. Medical Subject Headings (MESH) Fact Sheet. US National Library of Medicine. URL:http://www.nlm.nih.gov/pubs/factsheets/mesh.html (Accessed May 2011).

70. Cox RS, Long BC, Jones MI, Handler RJ. Sequestering of suffering: critical discourse analysis of natural disaster media coverage. Journal of Health Psychology. 2008;13:469-80.

71. Jones SC, Hall D, Cowlin F. Newspaper coverage of drug policy: an analysis of pre-election reporting of the Greens' drug policy in Australia. Drug and Alcohol Review. 2008;27:5-11.

72. Hulsen T, de Vlieg J, Alkema W. BioVenn - a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. BMC Genomics. 2008;9:488.

73. Dong D, Chang T-K, Chen D. Reporting AIDS and the invisible victims in China: official knowledge as news in the People's Daily, 1986-2002. Journal of health communication. 2008;13:357-74.

74. Abdelmutti N, Hoffman-Goetz L. Risk messages about HPV, cervical cancer, and the HPV vaccine Gardasil: a content analysis of Canadian and U.S. national newspaper articles. Women & health. 2009;49:422-40.

75. Forster A, Wardle J, Stephenson J, Waller J. Passport to promiscuity or lifesaver: press coverage of HPV vaccination and risky sexual behavior. Journal of health communication. 2010;15:205-17.

76. Hilton S, Hunt K, Langan M, Bedford H, Petticrew M. Newsprint media representations of the introduction of the HPV vaccination programme for cervical cancer prevention in the UK (2005-2008). Social science & medicine (1982). 2010;70:942-50.

77. Kelly BJ, Leader AE, Mittermaier DJ, Hornik RC, Cappella JN. The HPV vaccine and the media: how has the topic been covered and what are the effects on knowledge about the virus and cervical cancer? Patient Education and Counseling. 2009;77:308-13.

78. Smith MJ, Ellenberg SS, Bell LM, Rubin DM. Media coverage of the measles-mumps-rubella vaccine and autism controversy and its relationship to MMR immunization rates in the United States. Pediatrics. 2008;121:e836-43.

79. Duncan B. How the media reported the first days of the pandemic (H1N1) 2009: results of EU-wide media analysis. Eurosurveillance. 2009;14:19286.

80. Cohen EL, Caburnay CA, Luke DA, Rodgers S, Cameron GT, Kreuter MW. Cancer coverage in general-audience and Black newspapers. Health communication. 2008;23:427-35.

81. McDonnell DD, Lee H-J, Kim Y-B, Kazinets G, Moskowitz JM. Cancer coverage in a mainstream and Korean American online newspaper: lessons for community intervention. Patient Education and Counseling. 2008;71:388-95.

82. Tonmyr L, Jack S. Media analysis of early dissemination of Canadian child maltreatment surveillance data. Child Welfare. 2010;89:81-101.

83. Coker EM. Claiming the public soul: representations of Qur'anic healing and psychiatry in the Egyptian print media. Transcultural Psychiatry. 2009;46:672-94.

84. Feeley TH, Vincent D. How organ donation is represented in newspaper articles in the United States. Health communication. 2007;21:125-31.

85. MacKenzie R, Imison M, Chapman S, Holding S. Mixed messages and a missed opportunity: Australian news media coverage of Clare Oliver's campaign against solaria. The Medical Journal of Australia. 2008;189:371-4.

86. Moriarty CM, Stryker JE. Prevention and screening efficacy messages in newspaper accounts of cancer. Health Education Research. 2008;23:487-98.

87. Slater MD, Long M, Bettinghaus EP, Reineke JB. News coverage of cancer in the United States: a national sample of newspapers, television, and magazines. Journal of health communication. 2008;13:523-37.

88. Stryker JE, Emmons KM, Viswanath K. Uncovering differences across the cancer control continuum: a comparison of ethnic and mainstream cancer newspaper stories. Preventive Medicine. 2007;44:20-5.

89. Wilson M. Readability and patient education materials used for low-income populations. Clinical Nurse Specialist. 2009;23:33-40; quiz 41-2.

90. Racine E, Waldman S, Rosenberg J, Illes J. Contemporary neuroscience in the media. Social science & medicine (1982). 2010;71:725-33.

91. Niederdeppe J, Farrelly MC, Wenter D. Media advocacy, tobacco control policy change and teen smoking in Florida. Tobacco control. 2007;16:47-52.

92. Scully M, Wakefield M, Dixon H. Trends in news coverage about skin cancer prevention, 1993-2006: increasingly mixed messages for the public. Australian and New Zealand Journal of Public Health. 2008;32:461-6.

93. Quick BL. Applying the health belief model to examine news coverage regarding steroids in sports by ABC, CBS, and NBC between March 1990 and May 2008. Health communication. 2010;25:247-57.

94. Niederkrotenthaler T, Till B, Kapusta ND, Voracek M, Dervic K, Sonneck G. Copycat effects after media reports on suicide: a population-based ecologic study. Social science & medicine (1982). 2009;69:1085-90.

95. Belanger M, Harris PG, Nikolis A, Danino AM. Comparative analysis between scientific and the media communication following facial transplantation. Transplantation proceedings. 2009;41:485-8.

96. Payne BK, Appel J, Kim-Appel D. Elder abuse coverage in newspapers: regional differences and its comparison to child-abuse coverage. Journal of Elder Abuse & Neglect. 2008;20:265-75.

97. O'Hara SK, Smith KC. Presentation of eating disorders in the news media: What are the implications for patient diagnosis and treatment? Patient Education and Counseling. 2007;68:43-51.

98. Hong T. Information control in time of crisis: the framing of SARS in China-based newspapers and Internet sources. Cyberpsychology & Behavior. 2007;10:696-9.

99. Cohen T, Widdows D. Empirical distributional semantics: Methods and biomedical applications. Journal of Biomedical Informatics. 2009;42:390-405.

100. Cai J, Yang L, Liu Z, Ma Z, Liu Y. Comprehensive analysis of cancer coverage in important Chinese newspapers between 2000 and 2007. Support Care Cancer. 2009;17:329-32.

101. Fu KW, Yip PSF. Changes in reporting of suicide news after the promotion of the WHO media recommendations. Suicide & life-threatening behavior. 2008;38:631-6.

102. Gould MS, Midle JB, Insel B, Kleinman M. Suicide reporting content analysis: abstract development and reliability. Crisis. 2007;28:165-74.

103. Hagihara A, Tarumi K, Abe T. Media suicide-reports, Internet use and the occurrence of suicides between 1987 and 2005 in Japan. BMC Public Health. 2007;7:321.

104. Niederkrotenthaler T, Till B, Herberth A, Voracek M, Kapusta ND, Etzersdorfer E, Strauss M, Sonneck G. The gap between suicide characteristics in the print media and in the population. European Journal of Public Health. 2009;19:361-4.

105. Pirkis J, Burgess P, Blood RW, Francis C. The newsworthiness of suicide. Suicide & life-threatening behavior. 2007;37:278-83.

106. Dasgupta N, Mandl KD, Brownstein JS. Breaking the news or fueling the epidemic? Temporal association between news media report volume and opioid-related mortality. PLoS ONE. 2009;4:e7758.

107. Schwartz J, Andsager JL. Sexual health and stigma in urban newspaper coverage of methamphetamine. American Journal of Men's Health. 2008;2:57-67.

108. Weidner RR. Methamphetamine in three small Midwestern cities: evidence of a moral panic. Journal of Psychoactive Drugs. 2009;41:227-39.

109. Bengs C, Johansson E, Danielsson U, Lehti A, Hammarström A. Gendered portraits of depression in Swedish newspapers. Qualitative Health Research. 2008;18:962-73.

110. Gough B. 'Real men don't diet': an analysis of contemporary newspaper representations of men, food and health. Social science & medicine (1982). 2007;64:326-37.

111. Entman RM. Framing bias: Media in the distribution of power. Journal of Communication. 2007;57:163-173.

112. Basu AJ, Hogard E. Fit for public consumption? An exploratory study of the reporting of nutrition research in UK tabloids with regard to its accuracy, and a preliminary investigation of public attitudes towards it. Public Health Nutritian. 2008;11:1124-31.

113. Akamatsu R, Naito M, Nakayama T. Characteristics of reporting diabetes mellitus research results in Japanese newspapers. Bioscience Trends. 2009;3:44-7.

114. Berry TR, Wharf-Higgins J, Naylor PJ. SARS wars: an examination of the quantity and construction of health information in the news media. Health communication. 2007;21:35-44.

115. Brown L, Parker M, Dixon-Woods M. Whose interest? British newspaper reporting of use of medical records for research. Journal of Health Services Research & Policy. 2008;13:140-5.

116. Boke O, Aker S, Alptekin Aker A, Sarisoy G, Sahin AR. Schizophrenia in Turkish newspapers : retrospective scanning study. Social psychiatry and psychiatric epidemiology. 2007;42:457-61.

117. Clement S, Foster N. Newspaper reporting on schizophrenia: a content analysis of five national newspapers at two time points. Schizophr Res. 2008;98:178-83.

118. Carpiniello B, Girau R, Orrù MG. Mass-media, violence and mental illness. Evidence from some Italian newspapers. Epidemiologia e Psichiatria Sociale. 2007;16:251-5.

119. Wilson K, Code C, Dornan C, Ahmad N, Hébert P, Graham I. The reporting of theoretical health risks by the media: Canadian newspaper reporting of potential blood transmission of Creutzfeldt-Jakob disease. BMC Public Health. 2004;4:1.

120. Wilson K, Code C, Dornan C, Ahmad N, Hebert P, Graham I. The reporting of theoretical health risks by the media: Canadian newspaper reporting of potential blood transmission of Creutzfeldt-Jakob disease. BMC Public Health. 2004;4:1.

121. Theodoridis S, Koutroumbas K. Linear Classifiers. In: Pattern recognition. San Diego, CA: Elsevier; 2006:837.

122. Boehm B. A spiral model of software development and enhancement. Computer. 1988;21:61 - 72.

123. Agile Alliance. Manifesto for Agile Software Development. 2000. URL: http://www.agilemanifesto.org/ (Accessed February 3, 2011)

124. Larman C, Basili VR. Iterative and Incremental Development: A Brief History. IEEE Computer Society. 2003;36:47-56.

125. Amazon.com. What are Statistically Improbable Phrases. Journal. 2010:

126. Seale C, Ziebland S, Charteris-Black J. Gender, cancer experience and internet use: a comparative keyword analysis of interviews and online cancer support groups. Social science & medicine (1982). 2006;62:2577-90.

127. Salton G, Buckley C. Term-weighting approaches in automatic text retrieval. Information processing & management. 1988;24:513-523.

128. Chen SF, Goodman J. An Empirical Study of Smoothing Techniques for Language Modeling. Technical Report TR-10-98. 1998, Harvard University Center for Research in Computing Technology

129. Elhadad N. User-Sensitive Text Summarization: Application to the Medical Domain. Ph.D. Thesis, 2006.

130. Kullback S, Leibler R. On information and sufficiency. The Annals of Mathematical Statistics. 1951;22:79-86.

131. McFarlane DJ, Elhadad N, Kukafka R. Perplexity analysis of obesity news coverage. AMIA Annual Symposium Proceedings. 2009:426-30.

132. Deerwester S, Dumais S, Furnas G, Landauer T, Harshman R. Indexing by latent semantic analysis. Journal of the American society for information science. 1990;41:

133. Hofmann T. Probabilistic Latent Semantic Indexing. 2000;

134. McCallum A. Multi-label text classification with a mixture model trained by EM. AAAI workshop on Text Learning. 1999;

135. Sahlgren M. An introduction to random indexing. Methods and Applications of Semantic Indexing Workshop at the 7th International Conference on Terminology and Knowledge Engineering, TKE 2005. 2005;

136. Blei D, Ng A, Jordan M. Latent dirichlet allocation. The Journal of Machine Learning Research. 2003;3:

137. Geman S, Geman D. Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. IEEE Transactions on Pattern Analysis and Machine Intelligence. 1984;PAMI-6:721-741.

138. Wei X, Croft W. LDA-based document models for ad-hoc retrieval. SIGIR '06: Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval. 2006;

139. Dreyer L. Readability and responsibility. Journal of Reading. 1984;27:334-338.

140. Gemoets D, Rosemblat G, Tse T, Logan R. Assessing readability of consumer health information: an exploratory study. Stud Health Technol Inform. 2004;107:869-73.

141. Hyun S, Johnson SB, Bakken S. Exploring the ability of natural language processing to extract data from nursing narratives. Comput Inform Nurs. 2009;27:215-23; quiz 224-5.

142. Schwartz AS, Hearst MA. A simple algorithm for identifying abbreviation definitions in biomedical text. Pac Symp Biocomput. 2003;451-62.

143. Sheppard JE, Weidner LCE, Zakai S, Fountain-Polley S, Williams J. Ambiguous abbreviations: an audit of abbreviations in paediatric note keeping. Archives of Disease in Childhood. 2008;93:204-6.

144. Meade C, Smith C. Readability formulas: cautions and criteria. Patient Education and Counseling. 1991;

145. Flesch R. A new readability yardstick. Journal of Applied Psychology. 1948;32:221-33.

146. Smith E, Kincaid J. Derivation and validation of the automated readability index for use with technical materials. Human Factors: The Journal of the Human Factors and Ergonomics Society. 1970;12:457-464.

147. Mclaughlin GH. SMOG Grading-a New Readability Formula. Journal of Reading. 1969;12:639-646.

148. Berland GK, Elliott MN, Morales LS, Algazy JI, Kravitz RL, Broder MS, Kanouse DE, Muñoz JA, Puyol JA, Lara M, Watkins KE, Yang H, McGlynn EA. Health information on the Internet: accessibility, quality, and readability in English and Spanish. JAMA. 2001;285:2612-21.

149. Williamson JML, Martin AG. Assessing the readability statistics of national consent forms in the UK. International Journal of Clinical Practice. 2010;64:322-9.

150. MacKenzie R, Chapman S, Holding S, McGeechan K. 'A matter of faith, not science': analysis of media coverage of prostate cancer screening in Australian news media 2003-2006. Journal of the Royal Society of Medicine. 2007;100:513-21.

151. Slater MD, Long M, Ford VL. Alcohol, illegal drugs, violent crime, and traffic-related and other unintended injuries in U.S. local and national news. Journal of Studies on Alcohol. 2006;67:904-10.

152. Boyd AD, Jardine CG, Driedger SM. Canadian media representations of mad cow disease. Journal of Toxicology and Environmental Health. 2009;72:1096-105.

153. Pang B, Lee L, Vaithyanathan S. Thumbs up?: sentiment classification using machine learning techniques. EMNLP '02: Proceedings of the ACL-02 conference on Empirical methods in natural language processing. 2002;10:79 - 86.

154. Whissell C. Dictionary of Affect in Language. Journal. 1989;

155. Godbole N, Srinivasaiah M, Skiena S. Large-scale sentiment analysis for news and blogs. Proceedings of the International Conference on Weblogs and Social Media (ICWSM). 2007;

156. Pennebaker JW, Chung CK, Ireland M, Gonzales A, Booth RJ. 2007, Linguistic Inquiry and Word Count. URL:http://www.liwc.net (Accessed Feb 2010).

157. Theodoridis S, Koutroumbas K. Classifiers Based On Bayes Decision Theory. In: Pattern recognition. San Diego, CA: Elsevier; 2006:837.

158. Wagstaff K, Cardie C, Rogers S, Schrödl S. Constrained K-means Clustering with Background Knowledge. ICML '01: Proceedings of the Eighteenth International Conference on Machine Learning. 2001;

159. Blum A, Mitchell T. Combining labeled and unlabeled data with co-training. COLT' 98: Proceedings of the eleventh annual conference on Computational learning theory. 1998;

160. Theodoridis S, Koutroumbas K. Clustering Algorithms III. In: Pattern recognition. San Diego, CA: Elsevier; 2006:837.

161. Weston P. K Means Example Step 1-4. 2007. URL: http://en.wikipedia.org/wiki/K-means_clustering (Accessed 7/29/2010)

162. Hripcsak G, Wilcox A. Reference standards, judges, and comparison subjects: roles for experts in evaluating system performance. Journal of the American Medical Informatics Association : JAMIA. 2002;9:1-15.

163. Foley D. Considerations of sample and feature size. IEEE Transactions on Information Theory. 1972;18: 618- 626.

164. Taibi F, Abbou F, Alam M. A matching approach for object-oriented formal specifications. Journal of Object Technology. 2008;7:139-153.

165. Xiong J, Rayner S, Luo K, Li Y, Chen S. Genome wide prediction of protein function via a generic knowledge discovery approach based on evidence integration. BMC Bioinformatics. 2006;7:268.

166. Smucker M, Allan J, Carterette B. A comparison of statistical significance tests for information retrieval evaluation. CIKM '07: Proceedings of the sixteenth ACM conference on Conference on information and knowledge management. 2007;

167. Yeh A. More accurate tests for the statistical significance of result differences. Proceedings of the 18th conference on Computational linguistics. 2000;2:

168. Theodoridis S, Koutroumbas K. Feature Selection. In: Pattern recognition. San Diego, CA: Elsevier; 2006:213-262.

169. Fawcett T. An introduction to ROC analysis. Pattern recognition letters. 2006;27:861-874.

170. Schapire R, Singer Y. BoosTexter: A boosting-based system for text categorization. Machine Learning. 2000;39:135-168.

171. Meneghin G, Kreiner J. The Boosting Approach to Machine Learning An Overview (Robert E. Schapire). 2009;

172. McFarlane DJ, Kukafka R. Using term frequency to identify trends in the media's coverage of health. AMIA Annual Symposium Proceedings. 2007:1045.

173. Graff D. English Gigaword. Linguistic Data Consortium, Philadelphia. 2003.

174. Alias-i. 2010, LingPipe. 4.0.0. URL:http://alias-i.com/lingpipe (Accessed December 2010).

175. Stolcke A. SRILM-an Extensible Language Modeling Toolkit. Seventh International Conference on Spoken Language …. 2002;

176. Nichols S, Chase N. The quality of health research reporting by the daily newspapers in Trinidad and Tobago. The West Indian medical journal. 2005;54:302-7.

177. Noto AR, Pinsky I, Mastroianni FdeC. Drugs in the Brazilian print media: an exploratory survey of newspaper and magazine stories in the year 2000. Substance use & misuse. 2006;41:1263-76.

178. Sing T, Sander O, Beerenwinkel N, Lengauer T. ROCR: visualizing classifier performance in R. Bioinformatics. 2005;21:3940-1.

179. Aarva P. Studying the striving and opposing forces in newspaper journalism: the actantial model of health promotion. Health Promotion International. 2006;21:160-168.

180. Cohen J. A coefficient of agreement for nominal scales. Educational and psychological measurement. 1960;20:37-46.

181. M.F.Porter. An algorithm for suffix stripping. Program: electronic library and information systems. 1980;14:130-137.

182. Calais. ClearForest. URL:http://www.opencalais.com (Accessed October 15, 2008).

183. Phan X-H, Nguyen L-M, Horiguchi S. Learning to classify short and sparse text & web with hidden topics from large-scale data collections. WWW '08: Proceeding of the 17th international conference on World Wide Web. 2008;

184. Phan X-H, Nguyen C-T. GibbsLDA++. Journal. 2007;2010:

185. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten I. The WEKA data mining software: an update. SIGKDD Explorations Newsletter. 2009;11:

186. Zhang W, Wang H, Song SW, Fuller GN. Insulin-like growth factor binding protein 2: gene expression microarrays and the hypothesis-generation paradigm. Brain Pathology. 2002;12:87-94.

187. Fisher P, Hedeler C, Wolstencroft K, Hulme H, Noyes H, Kemp S, Stevens R, Brass A. A systematic strategy for large-scale analysis of genotype phenotype correlations: identification of candidate genes involved in African trypanosomiasis. Nucleic Acids Research. 2007;35:5625-33.

188. Khiabanian H, Holmes AB, Kelly BJ, Gururaj M, Hripcsak G, Rabadan R. Signs of the 2009 influenza pandemic in the New York-Presbyterian Hospital electronic health records. PLoS ONE. 2010;5:

189. Garfield E. Linking literatures: An intriguing use of the citation index. Current Contents. (May 23, 1994);3-5.

190. Swanson DR. Fish oil, Raynaud's syndrome, and undiscovered public knowledge. Perspectives in biology and medicine. 1986;30:7-18.

191. DiGiacomo RA, Kremer JM, Shah DM. Fish-oil dietary supplementation in patients with Raynaud's phenomenon: a double-blind, controlled, prospective study. Am J Med. 1989;86:158-64.

192. Jensen LJ, Saric J, Bork P. Literature mining for the biologist: from information retrieval to biological discovery. Nat Rev Genet. 2006;7:119-29.

193. Cohen AM, Hersh WR. A survey of current work in biomedical text mining. Briefings in Bioinformatics. 2005;6:57-71.

194. Zweigenbaum P, Demner-Fushman D, Yu H, Cohen KB. Frontiers of biomedical text mining: current progress. Briefings in Bioinformatics. 2007;8:358-75.

195. Google.org Flu Trends. Google.org. URL:http://www.google.org/flutrends/ (Accessed January 15, 2011).

196. Freifeld CC, Mandl KD, Reis BY, Brownstein JS. HealthMap: global infectious disease monitoring through automated classification and visualization of Internet media reports. Journal of the American Medical Informatics Association : JAMIA. 2008;15:150-7.

197. Stryker JE, Fishman J, Emmons KM, Viswanath K. Cancer risk communication in mainstream and ethnic newspapers. Preventing Chronic Disease. 2009;6:A23.

198. Amazon Mechanical Turk. Journal.

199. Barzilay R, Lee L. Catching the drift: Probabilistic content models, with applications to generation and summarization. Proceedings of HLT-NAACL 2004. 2004;

200. Augoustinos M, Crabb S, Shepherd R. Genetically modified food in the news: media representations of the GM debate in the UK. Public Understanding of Science. 2010;19:98-114.

201. Balasegaram M, Balasegaram S, Malvy D, Millet P. Neglected diseases in the news: a content analysis of recent international media coverage focussing on leishmaniasis and trypanosomiasis. PLoS Neglected Tropical Diseases. 2008;2:e234.

202. Barry CL, Busch SH. News coverage of FDA warnings on pediatric antidepressant use and suicidality. Pediatrics. 2010;125:88-95.

203. Barss P, Subait OM, Ali MHA, Grivna M. Drowning in a high-income developing country in the Middle East: newspapers as an essential resource for injury surveillance. Journal Science and Medicine in Sport. 2009;12:164-70.

204. Bartley M, Tobin WO, Collins R, O'Neill D. Stroke and newspapers: inattention or neglect? Journal of Stroke Cerebrovascular Diseases. 2009;18:259-61.

205. Boholm M. Risk and causality in newspaper reporting. Risk Analysis. 2009;29:1566-77.

206. Bonevski B, Wilson A, Henry DA. An analysis of news media coverage of complementary and alternative medicine. PLoS ONE. 2008;3:e2406.

207. Boyce T, Murray E, Holmes A. What are the drivers of the UK media coverage of meticillin-resistant Staphylococcus aureus, the inter-relationships and relative influences? The Journal of Hospital Infection. 2009;73:400-7.

208. Brown C. The information trail of the 'Freshman 15'--a systematic review of a health myth within the research and popular literature. Health Information & Libraries Journal. 2008;25:1-12.

209. Canales MK, Breslau ES, Nelson DE, Ballard-Barbash RR. Did news reporters get it right? Translation of the 2002 hormone study findings. American Journal of preventive medicine. 2008;34:61-8.

210. Caulfield T, Bubela T, Murdoch CJ. Myriad and the mass media: the covering of a gene patent controversy. Genetics in Medicine. 2007;9:850-5.

211. Chan P, Dipper A, Kelsey P, Harrison J. Newspaper reporting of meticillin-resistant Staphylococcus aureus and 'the dirty hospital'. The Journal of Hospital Infection. 2010;75:318-22.

212. Chopra AK, Doody GA. Schizophrenia, an illness and a metaphor: analysis of the use of the term 'schizophrenia' in the UK national newspapers. Journal of the Royal Society of Medicine. 2007;100:423-6.

213. Colak MY, Hekimoglu D, Ersoy K, Sozen F, Haberal M. Health content analysis of organ donation and transplantation news on Turkish television channels and in Turkish print media. Transplantation proceedings. 2010;42:134-6.

214. De Jonge J, Van Trijp H, Renes RJ, Frewer LJ. Consumer confidence in the safety of food and newspaper coverage of food safety issues: a longitudinal perspective. Risk Analysis. 2010;30:125-42.

215. Driedger SM. Risk and the media: a comparison of print and televised news stories of a Canadian drinking water risk event. Risk Analysis. 2007;27:775-86.

216. Euba R, Crugel M. The depiction of electroconvulsive therapy in the British press. The Journal of ECT. 2009;25:265-9.

217. Gollust SE, Lantz PM. Communicating population health: Print news media coverage of type 2 diabetes. Social science & medicine (1982). 2009;69:1091-1098.

218. Gross DP, Field J, Shanski K, Ferrari R. An evaluation of low back-pain-related content in Canadian newspaper media. Journal of spinal disorders & techniques. 2008;21:1-3.

219. Hahn W, Morley CP, Morrow C, Epling JW. The effect of media attention on concern for and medical management of methicillin-resistant Staphylococcus aureus: a multimethod study. Journal of Public Health Management and Practice. 2009;15:150-9.

220. Harris JK, Shelton SC, Moreland-Russell S, Luke DA. Tobacco coverage in print media: the use of timing and themes by tobacco control supporters and opposition before a failed tobacco tax initiative. Tobacco control. 2010;19:37-43.

221. Hayes M, Ross IE, Gasher M, Gutstein D, Dunn JR, Hackett RA. Telling stories: news media, health literacy and public policy in Canada. Social science & medicine (1982). 2007;64:1842-52.

222. Heneghan MK, Hazan C, Halpern AC, Oliveria SA. Skin cancer coverage in a national newspaper: a teachable moment. Journal of Cancer Education. 2007;22:99-104.

223. Hill PS, Ngo AD, Khuong TA, Dao HL, Hoang HTM, Trinh HT, Nguyen LTN, Nguyen PH. Mandatory helmet legislation and the print media in Viet Nam. Accident, Analysis and Prevention. 2009;41:789-97.

224. Hilliard N, Jenkins R, Pashayan N, Powles J. Informal knowledge transfer in the period before formal health education programmes: case studies of mass media coverage of HIV and SIDS in England and Wales. BMC Public Health. 2007;7:293.

225. Hilliker L. The reporting of grief by one newspaper of record for the U.S.: the New York Times. Omega (Westport). 2008;57:261-78.

226. Hochman M, Hochman S, Bor D, McCormick D. News media coverage of medication research: reporting pharmaceutical company funding and use of generic medication names. JAMA. 2008;300:1544-50.

227. Jensen JD, Moriarty CM, Hurley RJ, Stryker JE. Making sense of cancer news coverage trends: a comparison of three comprehensive content analyses. Journal of health communication. 2010;15:136-51.

228. Kisely S. Surveying the portrayal of mental illness across a jurisdiction: is more than one method appropriate? Public Health. 2008;122:506-8.

229. Kromm EE, Smith KC, Singer RF. Survivors on Cancer: the portrayal of survivors in print news. Journal of cancer survivorship : research and practice. 2007;1:298-305.

230. Lai WYY, Lane T, Jones A. Sources and coverage of medical news on front pages of US newspapers. PLoS ONE. 2009;4:e6856.

231. Lai WYY, Lane T. Characteristics of medical research news reported on front pages of newspapers. PLoS ONE. 2009;4:e6103.

232. Larson S, Long M, Slater MD, Bettinghaus EP, Read A. A content analysis of cancer survivorship coverage in a representative sample of US news outlets. Journal of Cancer Education. 2009;24:291-6.

233. Lawrence J, Kearns RA, Park J, Bryder L, Worth H. Discourses of disease: representations of tuberculosis within New Zealand newspapers 2002-2004. Social science & medicine (1982). 2008;66:727-39.

234. Li M, Chapman S, Agho K, Eastman CJ. Can even minimal news coverage influence consumer health-related behaviour? A case study of iodized salt sales, Australia. Health Education Research. 2008;23:543-8.

235. Løchen M-L, Gram IT, Skattebo S, Kolstrup N. Tobacco images and texts in Norwegian magazines and newspapers. Scand J Public Health. 2007;35:31-8.

236. Mack N, Robinson ET, Macqueen KM, Moffett J, Johnson LM. The exploitation of "Exploitation" in the tenofovir prep trial in Cameroon: Lessons learned from media coverage of an HIV prevention trial. Journal of Empirical Research on Human Research Ethics. 2010;5:3-19.

237. MacKenzie R, Chapman S, Barratt A, Holding S. "The news is [not] all good": misrepresentations and inaccuracies in Australian news media reports on prostate cancer screening. The Medical Journal of Australia. 2007;187:507-10.

238. Matamoros DJC, Axelsson R, Strid J. How do newspapers deal with health in Sweden? A descriptive study. Patient Education and Counseling. 2007;67:78-83.

239. Mcleod K, Wakefield M, Chapman S, Smith KC, Durkin S. Changes in the news representation of smokers and tobacco-related media advocacy from 1995 to 2005 in Australia. Journal of epidemiology and community health. 2009;63:215-20.

240. Mizuno Y, Narimatsu H, Kishi Y, Kodama Y, Murashige N, Yuji K, Matsumura T, Kami M. Structural problems of medical news reports in newspapers: a verification of news reports on an incident of mass nosocomial Serratia infection. Journal of Infection and Chemotherapy. 2010;16:107-12.

241. Nelson DE, Signorielli N. Reporter sex and newspaper coverage of the adverse health effects of hormone therapy. Women & health. 2007;45:1-15.

242. Nelson MJ, DeIorio NM, Schmidt T, Griffiths D, Daya M, Haywood L, Zive D, Newgard CD. Local media influence on opting out from an exception from informed consent trial. Annals of Emergency Medicine. 2010;55:1-8.

243. Niederdeppe J, Frosch DL, Hornik RC. Cancer news coverage and information seeking. Journal of health communication. 2008;13:181-99.

244. Niederdeppe J, Frosch DL. News coverage and sales of products with trans fat: effects before and after changes in federal labeling policy. American Journal of preventive medicine. 2009;36:395-401.

245. Rabi DM, Lewin AM, Brown GE, Edwards AL, Johnson JA, Ghali WA. Lay media reporting of rosiglitazone risk: extent, messaging and quality of reporting. Cardiovascular Diabetology. 2009;8:40.

246. Racine E, Amaram R, Seidler M, Karczewska M, Illes J. Media coverage of the persistent vegetative state and end-of-life decision-making. Neurology. 2008;71:1027-32.

247. Rada RF. Retractions, press releases and newspaper coverage. Health Information & Libraries Journal. 2007;24:210-5.

248. Raman U. Medical ethics in the media. Indian journal of medical ethics. 2009;6:19-24.

249. Reid AJ, Malone PSC. Plastic surgery in the press. Journal of plastic, reconstructive & aesthetic surgery : JPRAS. 2008;61:866-9.

250. Rosales M, Stallones L. Coverage of motor vehicle crashes with injuries in U.S. newspapers, 1999-2002. Journal of Safety Research. 2008;39:477-82.

251. Seale C, Boden S, Williams S, Lowe P, Steinberg D. Media constructions of sleep and sleep disorders: a study of UK national newspapers. Social science & medicine (1982). 2007;65:418-30.

252. Slopen NB, Watson AC, Gracia G, Corrigan PW. Age analysis of newspaper coverage of mental illness. Journal of health communication. 2007;12:3-15.

253. Smith KC, Wakefield MA, Terry-Mcelrath Y, Chaloupka FJ, Flay B, Johnston L, Saba A, Siebel C. Relation between newspaper coverage of tobacco issues and smoking attitudes and behaviour among American teens. Tobacco control. 2008;17:17-24.

254. Stephenson N, Jamieson M. Securitising health: Australian newspaper coverage of pandemic influenza. Sociology of Health & Illness. 2009;31:525-39.

255. Stryker JE, Moriarty CM, Jensen JD. Effects of newspaper coverage on public knowledge about modifiable cancer risks. Health communication. 2008;23:380-90.

256. Takahashi Y, Miyaki K, Nakayama T. Analysis of news of the Japanese asbestos panic: a supposedly resolved issue that turned out to be a time bomb. Journal of public health (Oxford, England). 2007;29:62-9.

257. Tong A, Chapman S, Sainsbury P, Craig JC. An analysis of media coverage on the prevention and early detection of CKD in Australia. Am J Kidney Dis. 2008;52:159-70.

258. Wilson PM, Booth AM, Eastwood A, Watt IS. Deconstructing media coverage of trastuzumab (Herceptin): an analysis of national newspaper coverage. Journal of the Royal Society of Medicine. 2008;101:125-32.

259. Yong PL, Bigman C, Flynn DN, Mittermaier D, Long JA. Messages about black-box warnings: a comparative analysis of reports from the FDA and lay media in the US. Drug Safety. 2009;32:1147-57.

APPENDIX A. LITERATURE REVIEW RESEARCH ARTICLES

Variable	**Value**	**Description**
Media effects	AGS	Agenda setting
	FR	Argument framing
	INF	Informing
IR methods	keyword	Keyword search
	sample	News sampling
	manual	Manual filtering
	pre	Pre-categorized news
	clipping	Third party clipping service
Coding method	CA	Manual content analysis
	keyword	Use of keyword appearance for coding
	auto	Automated or computer aided coding methodology
Analysis method	A	ANOVA (analysis of variance)
	χ^2	Chi-square test (Pearson's)
	F	Fisher's exact test
	γ	Gamma test
	κ	Cohen's kappa statistic
	L	Linear modeling
	M	McNemar test
	O	Odds ratios
	P	Principle components analysis
	Q	Cochran's Q test
	R	Regression analysis
	t	Student's *t*-test
	TS	Time series analysis
	U	Mann-Whitney U test / Wilcoxon rank-sum test
	Z	Z-test
	#	Solely descriptive analyses and comparisons of frequencies
Finding of news coverage quality	+	Overall good news quality
	-	Bad news quality
	+ / -	Mixed news quality
Topic Category		Refer to Table 4 on page 35

Note 1: *N/A* = The research article did not provide this information.
Note 2: For *analysis method* **all** studies report frequency counts and most compare raw frequencies. The "#" code is used when no analysis other than basic frequency comparisons was done.
Note 3: Research articles labeled using kappa for analysis indicate that the kappa, and the associated coder reliability was a finding of the research, as opposed to its being used solely to measure the level of agreement for coded variables that were in turn analyzed to produce findings.

Author	Topic Research Category Episodic news coverage	Corpus Size Geography Population	Methods	News Quality
Abdelmutti[74]	*Topic* = HPV *Category* = Prevention **Episodic**	*N* = 167 *Geo* = Canada & US *Pop* = Women, adolescent girls	*IR* = keyword *Type* = AGS FR *Coding* = CA *Analysis* = A R χ^2	**-**
Akamatsu[113]	*Topic* = Diabetes *Category* = Research **Non-episodic**	*N* = 152 *Geo* = Japan *Pop = N/A*	*IR* = keyword *Type* = INF *Coding* = CA *Analysis* = χ^2	**-**
Amzel[28]	*Topic* = Health Disparities *Category* = Disparities **Non-episodic**	*N* = 1188 *Geo* = US *Pop* = Blacks, Hispanics, Asian American, Native American	*IR* = keyword *Type* = AGS *Coding* = CA *Analysis = #*	**-**
Anderson[48]	*Topic* = Bedbugs *Category* = Risk **Non-episodic**	*N* = 395 [*] *Geo* = US *Pop = N/A*	*IR* = keyword *Type* = AGS *Coding* = CA *Analysis = #*	**N/A**
Andersson[36]	*Topic* = Diabetes *Category* = Prevention **Episodic**	*N* = 2128 *Geo* = Sweden *Pop = N/A*	*IR* = manual *Type* = AGS FR *Coding* = keyword *Analysis = #*	**N/A**
Atkin[55]	*Topic* = Breast cancer *Category* = Illness **Non-episodic**	*N* = 231 *Geo* = US *Pop* = Women	*IR* = keyword *Type* = INF *Coding* = CA *Analysis = #*	**-**

[*] This is an estimation based on charts in the articles.

Author	Topic Research Category Episodic news coverage	Corpus Size Geography Population	Methods	News Quality
Augoustinos[200]	*Topic* = Genetically modified food *Category* = Policies&Inst **Episodic**	*N* = 690 *Geo* = UK *Pop = N/A*	*IR* = keyword *Type* = AGS FR *Coding* = CA *Analysis* = #	**N/A**
Balasegaram[201]	*Topic* = Neglected diseases (Leishmaniasis and Trypanosomiasis) *Category* = Illness **Non-episodic**	*N* = 113 *Geo* = Worldwide *Pop = N/A*	*IR* = keyword *Type* = AGS FR *Coding* = CA *Analysis* = #	**-**
Barnes[37]	*Topic* = Disaster management (Katrina) *Category* = Risk **Episodic**	*N* = 1590 *Geo* = US *Pop = N/A*	*IR* = keyword *Type* = AGS FR *Coding* = CA *Analysis* = #	**+/-**
Barry[202]	*Topic* = Pediatric antidepressant *Category* = Policies&Inst **Episodic**	*N* = 167 *Geo* = US *Pop* = Children	*IR* = keyword *Type* = AGS INF FR *Coding* = CA *Analysis* = R	**+/-**
Barss[203]	*Topic* = Injury surveillance (drowning) *Category* = Injury **Non-episodic**	*N* = 79 *Geo* = United Arab Emirates *Pop = N/A*	*IR* = pre *Type* = INF *Coding* = CA *Analysis* = #	**+**
Bartley[204]	*Topic* = Stroke *Category* = Illness **Non-episodic**	*N* = 2344 *Geo* = UK *Pop = N/A*	*IR* = keyword manual *Type* = AGS *Coding* = CA *Analysis* = #	**+/-**
Basu[112]	*Topic* = Nutrition Research *Category* = Research **Non-episodic**	*N* = 29 *Geo* = UK *Pop = N/A*	*IR* = manual *Type* = INF FR *Coding* = CA *Analysis* = #	**-**
Belanger[95]	*Topic* = Facial transplantation *Category* = Treatment **Episodic**	*N* = 102 *Geo* = US & France *Pop = N/A*	*IR* = keyword *Type* = FR *Coding* = CA *Analysis* = #	**N/A**

Author	Topic Research Category Episodic news coverage	Corpus Size Geography Population	Methods	News Quality
Bengs[109]	*Topic* = Depression *Category* = Illness **Non-episodic**	*N* = 26 *Geo* = Sweden *Pop* = Men vs. women	*IR* = keyword manual *Type* = FR *Coding* = CA *Analysis* = #	**N/A**
Berry[114]	*Topic* = *Risk*in health New reporting *Category* = General **Non-episodic**	*N* = 98 *Geo* = Canada *Pop = N/A*	*IR* = sample *Type* = FR *Coding* = CA *Analysis = t*	**-**
Boholm[205]	*Topic* = Causality and *Risk*in reporting *Category* = Risk **Non-episodic**	*N* = 166 *Geo* = Sweden *Pop = N/A*	*IR* = keyword manual *Type* = INF *Coding* = auto *Analysis* = #	**N/A**
Boke[116]	*Topic* = Schizophrenia *Category* = Illness **Non-episodic**	*N* = 878 *Geo* = Turkey *Pop = N/A*	*IR* = keyword *Type* = INF *Coding* = CA *Analysis* = χ^2	**-**
Bonevski[206]	*Topic* = Alternative medicine *Category* = Treatment **Non-episodic**	*N* = 222 *Geo* = Australia *Pop = N/A*	*IR* = manual *Type* = INF *Coding* = CA *Analysis* = A L TS	**-**
Boyce[207]	*Topic* = Meticillin-resistant Staphylococcus aureus *Category* = Illness **Non-episodic**	*N* = 2880 *Geo* = UK *Pop = N/A*	*IR* = keyword *Type* = AGS FR INF *Coding* = CA *Analysis* = #	**N/A**
Boyd[152]	*Topic* = Mad cow disease *Category* = Illness **Episodic**	*N* = 309 *Geo* = Canada *Pop = N/A*	*IR* = keyword *Type* = FR *Coding* = CA *Analysis* = #	**N/A**
Brown[208]	*Topic* = Obesity *Category* = Risk **Non-episodic**	*N* = 165 *Geo* = US *Pop = N/A*	*IR* = keyword manual *Type* = AGS INF FR *Coding* = CA *Analysis* = #	**+/-**

Author	Topic Research Category Episodic news coverage	Corpus Size Geography Population	Methods	News Quality
Brown[115]	*Topic* = Medical records *Category* = Policies&Inst **Non-episodic**	*N* = 80 *Geo* = UK *Pop = N/A*	*IR* = keyword manual *Type* = AGS FR *Coding* = CA *Analysis = #*	**-**
Cai[100]	*Topic* = Cancer *Category* = Illness **Non-episodic**	*N* = 7653 *Geo* = China *Pop = N/A*	*IR* = keyword *Type* = AGS FR *Coding* = keyword *Analysis* = L TS	**-**
Canales[209]	*Topic* = Hormone therapy *Category* = Research **Episodic**	*N* = 198 *Geo* = US *Pop = N/A*	*IR* = keyword *Type* = FR INF *Coding* = CA *Analysis = #*	**+**
Carpiniello[118]	*Topic* = Mental illness *Category* = Injury **Non-episodic**	*N* = 2279 *Geo* = Italy *Pop = N/A*	*IR = N/A* *Type* = AGS FR *Coding* = CA *Analysis* = F *t* A χ^2	**-**
Caulfield[210]	*Topic* = Gene patenting *Category* = Policies&Inst **Episodic**	*N* = 143 *Geo* = Australia, Canada, US, UK *Pop = N/A*	*IR* = keyword *Type* = FR AGS *Coding* = CA *Analysis = #*	**+/-**
Caulfield[30]	*Topic* = Medications (BiDil) *Category* = Disparities **Episodic**	*N* = 104 *Geo* = US *Pop* = Blacks	*IR* = keyword *Type* = FR *Coding* = CA *Analysis* = χ^2	**+**
Chan[211]	*Topic* = Staphylococcus *Category* = Risk **Episodic**	*N* = 2129 * *Geo* = UK *Pop = N/A*	*IR* = keyword *Type* = AGS FR INF *Coding* = CA *Analysis = #*	**N/A**

* Only applies to argument framing study

Author	Topic Research Category Episodic news coverage	Corpus Size Geography Population	Methods	News Quality
Chau[40]	*Topic* = Exercise, smoking, obesity *Category* = Prevention **Non-episodic**	*N* = 84136 *Geo* = Australia *Pop = N/A*	*IR* = keyword sample *Type* = AGS *Coding* = CA keyword *Analysis* = L TS	**+/-**
Chopra[212]	*Topic* = Words schizophrenia and cancer as metaphors *Category* = Illness **Non-episodic**	*N* = 600 *Geo* = UK *Pop = N/A*	*IR* = keyword *Type* = FR *Coding* = CA *Analysis* = χ^2	**-**
Clement[117]	*Topic* = Schizophrenia *Category* = Illness **Non-episodic**	*N* = 1196 *Geo* = UK *Pop = N/A*	*IR* = keyword *Type* = FR *Coding* = CA *Analysis* = R χ^2	**-**
Cohen[80]	*Topic* = Cancer *Category* = Illness **Non-episodic**	*N* = 5206 *Geo* = US *Pop* = Blacks	*IR* = keyword sample *Type* = AGS INF *Coding* = CA *Analysis* = χ^2	**+/-**
Coker[83]	*Topic* = Psychiatry & faith healing *Category* = Treatment **Non-episodic**	*N* = 107 *Geo* = Egypt *Pop = N/A*	*IR* = keyword *Type* = FR *Coding* = CA *Analysis = #*	**N/A**
Colak[213]	*Topic* = Organ donation *Category* = Other **Non-episodic**	*N* = 2449 *Geo* = Turkey *Pop = N/A*	*IR = N/A* *Type* = FR *Coding* = CA *Analysis = #*	**-**
Dasgupta[106]	*Topic* = Drug abuse (prescription) *Category* = Risk **Non-episodic**	*N* = 31 *Geo* = US *Pop = N/A*	*IR* = keyword manual *Type* = AGS *Coding* = CA keyword *Analysis* = TS	**N/A**

Author	Topic Research Category Episodic news coverage	Corpus Size Geography Population	Methods	News Quality
De Jonge[214]	*Topic* = Food safety *Category* = Risk **Non-episodic**	*N* = 2087 *Geo* = Netherlands *Pop = N/A*	*IR* = keyword *Type* = AGS INF *Coding* = CA *Analysis* = #	**N/A**
Dong[73]	*Topic* = HIV *Category* = Illness **Non-episodic**	*N* = 607440, 155 [*] *Geo* = China *Pop = N/A*	*IR* = keyword sample *Type* = AGS INF FR *Coding* = CA keyword *Analysis* = #	**-**
Driedger[215]	*Topic* = drinking water *Category* = Risk **Episodic**	*N* = 1238 *Geo* = CANADA *Pop = N/A*	*IR* = keyword *Type* = FR *Coding* = CA *Analysis* = #	**N/A**
Duncan[79]	*Topic* = Pandemic influenza *Category* = Illness **Episodic**	*N* = 3979 *Geo* = 31 European countries [†] *Pop = N/A*	*IR* = keyword *Type* = AGS FR INF *Coding* = CA *Analysis* = #	**N/A**
Euba[216]	*Topic* = Electroconvulsive therapy *Category* = Treatment **Non-episodic**	*N* = 348 *Geo* = Great Britain *Pop = N/A*	*IR* = keyword *Type* = AFR *Coding* = CA *Analysis* = O	**N/A**
Feeley[84]	*Topic* = Organ donation *Category* = Policies&Inst **Non-episodic**	*N* = 715 *Geo* = US *Pop = N/A*	*IR* = keyword *Type* = FR AGS *Coding* = CA *Analysis* = χ^2	**N/A**

[*] Agenda setting was done on the larger number, informing and argument framing studies performed on the smaller number.
[†] 27 European Union (EU) Member States plus the 4 European Free Trade Association (EFTA) countries (Iceland, Liechtenstein, Norway and Switzerland)

Author	Topic Research Category Episodic news coverage	Corpus Size Geography Population	Methods	News Quality
Fishman [57]	*Topic* = Cancer *Category* = Treatment **Non-episodic**	*N* = 436 *Geo* = US *Pop = N/A*	*IR* = sample clipping *Type* = FR INF *Coding* = CA *Analysis* = χ^2	**-**
Forster [75]	*Topic* = HPV vaccination *Category* = Prevention **Episodic**	*N* = 92 *Geo* = US *Pop* = Women, adolescent girls	*IR* = keyword *Type* = AGS FR *Coding* = CA *Analysis* = #	**+**
Fu [101]	*Topic* = Suicide *Category* = Injury **Episodic**	*N* = 5740 *Geo* = China (Hong Kong) *Pop = N/A*	*IR* = keyword *Type* = AGS FR *Coding* = CA *Analysis* = χ^2	**+**
Gollust [217]	*Topic* = Diabetes *Category* = Illness **Non-episodic**	*N* = 859 *Geo* = US *Pop = N/A*	*IR* = keyword manual *Type* = AGS INF *Coding* = CA *Analysis* = R	**-**
Gould[102]	*Topic* = Suicide *Category* = Injury **Non-episodic**	*N* = 151 *Geo* = US *Pop = N/A*	*IR* = sample pre *Type* = INF FR *Coding* = CA *Analysis* = κ	**-**
Gross[218]	*Topic* = Back pain *Category* = Illness **Non-episodic**	*N* = 129 *Geo* = Canada *Pop = N/A*	*IR* = keyword *Type* = AGS INF FR *Coding* = CA *Analysis* = #	**+**
Hagihara[103]	*Topic* = Suicide *Category* = Injury **Non-episodic**	*N = N/A* [*] *Geo* = Japan *Pop = N/A*	*IR* = keyword *Type* = AGS *Coding* = CA *Analysis* = L TS	**N/A**

[*] They estimate the number of circulated news articles on suicide, without telling the number of articles specifically analyzed in the study.

Author	**Topic** **Research Category** **Episodic news coverage**	**Corpus Size** **Geography** **Population**	**Methods**	**News Quality**
Hahn [219]	*Topic* = Methicillin-resistant Staphylococcus aureus *Category* = Illness **Episodic**	*N* = 6000 * *Geo* = US *Pop = N/A*	*IR* = keyword *Type* = AGS *Coding* = keyword *Analysis* = #	**N/A**
Halpin [41]	*Topic* = Cancer (prostate) *Category* = Illness **Non-episodic**	*N* = 817 *Geo* = Canada *Pop* = Men	*IR* = keyword *Type* = FR INF *Coding* = CA *Analysis* = #	**-**
Harris [220]	*Topic* = Smoking *Category* = Risk **Episodic**	*N* = 1263 *Geo* = US (Missouri) *Pop = N/A*	*IR* = pre *Type* = FR *Coding* = CA *Analysis* = A χ^2	**+/-**
Hayes[221]	*Topic* = Canadian health stories *Category* = General **Non-episodic**	*N* = 4732 *Geo* = Canada *Pop = N/A*	*IR* = keyword sample manual *Type* = AGS *Coding* = CA *Analysis* = #	**-**
Henderson [25]	*Topic* = Obesity (childhood) *Category* = Risk **Episodic**	*N* = 717 *Geo* = Australia *Pop* = Children	*IR* = keyword *Type* = AGS FR *Coding* = CA *Analysis* = #	**N/A**
Heneghan[222]	*Topic* = Skin cancer *Category* = Illness **Non-episodic**	*N* = 874 *Geo* = US *Pop = N/A*	*IR* = keyword *Type* = AGS *Coding* = CA *Analysis* = #	**+/-**
Hilbert [38]	*Topic* = Obesity *Category* = Risk **Non-episodic**	*N* = 222 *Geo* = Germany *Pop = N/A*	*IR* = keyword *Type* = INF FR *Coding* = CA *Analysis* = U χ^2	**+/-**

* An exact figure was not provided. Estimate based on graphs in the article.

Author	Topic Research Category Episodic news coverage	Corpus Size Geography Population	Methods	News Quality
Hill[223]	*Topic* = Helmet use policies *Category* = Policies&Inst **Episodic**	*N* = 721 *Geo* = Viet Nam *Pop = N/A*	*IR = N/A* *Type* = FR *Coding* = CA *Analysis* = #	**+**
Hilliard[224]	*Topic* = HIV/SIDS *Category* = Illness **Non-episodic**	*N* = 299 *Geo* = UK *Pop* = MSM	*IR* = manual keyword *Type* = INF AGS *Coding* = CA *Analysis* = #	**N/A**
Hilliker[225]	*Topic* = Grief/coping *Category = Psych* **Non-episodic**	*N* = 50 *Geo* = US *Pop = N/A*	*IR* = keyword *Type* = FR *Coding* = CA *Analysis* = #	**+/-**
Hilton [76]	*Topic* = HPV vaccination *Category* = Prevention **Episodic**	*N* = 344 *Geo* = UK *Pop* = Women, adolescent girls	*IR* = keyword manual *Type* = FR *Coding* = CA *Analysis* = χ^2	**+**
Hochman[226]	*Topic* = Medical Research *Category* = Research **Non-episodic**	*N* = 306 *Geo* = US *Pop = N/A*	*IR* = keyword *Type* = AGS FR *Coding* = CA *Analysis* = χ^2	**-**
Hong[98]	*Topic* = SARS *Category* = Illness **Non-episodic**	*N* = 727 *Geo* = China *Pop = N/A*	*IR* = sample keyword *Type* = FR *Coding* = CA *Analysis* = P	**-**
Jensen [227]	*Topic* = Cancer *Category* = Illness **Non-episodic**	*N* = 5327 *Geo* = US *Pop = N/A*	*IR* = keyword *Type* = AGS FR INF *Coding* = CA *Analysis* = #	**-**
Kelly [77]	*Topic* = HPV vaccination *Category* = Prevention **Episodic**	*N* = 321 *Geo* = US *Pop* = Women, adolescent girls	*IR* = keyword *Type* = AGS INF *Coding* = CA *Analysis* = χ^2	**+/-**

Author	Topic Research Category Episodic news coverage	Corpus Size Geography Population	Methods	News Quality
Kim[24]	*Topic* = Obesity *Category* = Risk **Non-episodic**	*N* = 300 *Geo* = US *Pop = N/A*	*IR* = keyword sample *Type* = AGS FR *Coding* = CA *Analysis* = #	**+**
Kim [31]	*Topic* = Racial and ethnic health disparities *Category* = Disparities **Non-episodic**	*N* = 3823 *Geo* = US *Pop* = Various ethnicities	*IR* = keyword manual *Type* = AGS INF FR *Coding* = CA *Analysis* = TS *t*	**-**
Kisely[228]	*Topic* = Mental Illness *Category* = Illness **Non-episodic**	*N* = 202 *Geo* = Canada *Pop = N/A*	*IR* = clipping *Type* = INF *Coding* = CA *Analysis* = #	**+/-**
Kromm[229]	*Topic* = Cancer survivors *Category* = Illness **Non-episodic**	*N* = 517 *Geo* = US *Pop = N/A*	*IR* = keyword *Type* = INF *Coding* = CA *Analysis* = #	**+/-**
Lai [230] *	*Topic* = General health news *Category* = General **Non-episodic**	*N* = 1630 *Geo* = Worldwide *Pop = N/A*	*IR* = pre manual *Type* = AGS *Coding* = CA *Analysis* = χ^2	**N/A**
Lai [231]	*Topic* = Medical Research *Category* = Research **Non-episodic**	*N* = 734 *Geo* = US *Pop = N/A*	*IR* = keyword manual *Type* = INF *Coding* = CA *Analysis* = O χ^2	**+/-**
Larson [232]	*Topic* = Cancer *Category* = Illness **Non-episodic**	*N* = 412 *Geo* = US *Pop = N/A*	*IR* = sample *Type* = INF *Coding* = CA *Analysis* = χ^2	**-**

* 50 high-circulation, English-language national and regional newspapers worldwide

Author	**Topic** **Research Category** **Episodic news coverage**	**Corpus Size** **Geography** **Population**	**Methods**	**News Quality**
Lawrence[233]	*Topic* = Tuberculosis *Category* = Illness **Non-episodic**	*N* = 120 *Geo* = New Zealand *Pop = N/A*	*IR* = keyword sample *Type* = FR *Coding* = CA *Analysis* = #	**-**
Li[234]	*Topic* = Diet (iodized salt) *Category* = Prevention **Episodic**	*N* = 39 *Geo* = Australia *Pop = N/A*	*IR* = keyword *Type* = AGS *Coding* = keyword *Analysis* = R	**N/A**
Løchen[235]	*Topic* = Smoking *Category* = risk **Non-episodic**	*N* = 610 *Geo* = Norway *Pop = N/A*	*IR* = manual *Type* = AGS *Coding* = CA *Analysis* = χ^2	**+/-**
Mack [236]	*Topic* = HIV *Category* = Research **Episodic**	*N* = 134 *Geo* = Worldwide *Pop = N/A*	*IR* = manual *Type* = FR *Coding* = CA *Analysis* = #	**-**
MacKenzie[150]	*Topic* = Prostate cancer screening *Category* = Diagnosis **Episodic**	*N* = 388 *Geo* = Australia *Pop = N/A*	*IR = N/A* *Type* = FR AGS *Coding* = CA *Analysis* = #	**-**
MacKenzie[237]	*Topic* = Prostate cancer screening *Category* = Diagnosis **Non-episodic**	*N* = 388 *Geo* = Australia *Pop = N/A*	*IR = N/A* *Type* = INF *Coding* = CA *Analysis* = #	**-**

Author	Topic Research Category Episodic news coverage	Corpus Size Geography Population	Methods	News Quality
MacKenzie[85]	*Topic* = Cancer (Melanoma) *Category* = Risk **Episodic**	*N* = 83 *Geo* = Australia *Pop = N/A*	*IR* = keyword *Type* = FR *Coding* = CA *Analysis* = #	**-**
Matamoros[238]	*Topic* = Health in Swedish newspapers *Category* = General **Non-episodic**	*N* = 9767 * *Geo* = Sweden *Pop = N/A*	*IR* = manual *Type* = AGS FR *Coding* = CA *Analysis* = #	**+/-**
McDonnell[81]	*Topic* = Cancer *Category* = Illness **Non-episodic**	*N* = 300 *Geo* = US *Pop* = Korean Americans	*IR* = keyword sample *Type* = AGS FR INF *Coding* = CA *Analysis* = χ^2	**N/A**
Mcleod [239]	*Topic* = Smoking *Category* = Risk **Non-episodic**	*N* = 618 *Geo* = Australia *Pop = N/A*	*IR* = keyword sample *Type* = FR *Coding* = CA *Analysis* = R	**+/-**
Mizuno [240]	*Topic* = Mass nosocomial serratia infection *Category* = Illness **Episodic**	*N* = 188 *Geo* = Japan *Pop = N/A*	*IR* = keyword *Type* = AGS INF *Coding* = CA *Analysis* = #	**+/-**
Moriarty[86]	*Topic* = Cancer *Category* = Prevention **Non-episodic**	*N* = 2448 *Geo* = US *Pop = N/A*	*IR* = keyword manual *Type* = INF *Coding* = CA *Analysis* = χ^2	**-**
Moriarty [58]	*Topic* = Cancer *Category* = Illness **Non-episodic**	*N* = 3656 *Geo* = US *Pop = N/A*	*IR* = keyword sample *Type* = INF FR *Coding* = CA *Analysis* = γ	**N/A**

* Pages in newspaper, not news articles

Author	**Topic** **Research Category** **Episodic news coverage**	**Corpus Size** **Geography** **Population**	**Methods**	**News Quality**
Nelson[52]	*Topic* = Tobacco monitoring *Category* = risk **Non-episodic**	*N* = 1280 *Geo* = US *Pop = N/A*	*IR* = keyword *Type* = AGS FR INF *Coding* = CA *Analysis* = #	**N/A**
Nelson[241]	*Topic* = Hormone therapy *Category* = treatment **Episodic**	*N* = 111 *Geo* = US *Pop* = Women	*IR* = keyword manual *Type* = AGS FR INF *Coding* = CA *Analysis* = F *t* χ^2	**-**
Nelson [242]	*Topic* = Clinical trials (Notices informing public of opt-out procedures) *Category* = Research **Episodic**	*N* = 33 *Geo* = US *Pop = N/A*	*IR* = keyword *Type* = INF FR *Coding* = CA *Analysis* = #	**-**
Niederdeppe[91]	*Topic* = Tobacco control policy *Category* = Policies&Inst **Episodic**	*N* = 2330 *Geo* = US *Pop* = Adolescents	*IR* = clipping *Type* = AGS *Coding* = CA *Analysis* = O R	**N/A**
Niederdeppe[243]	*Topic* = Cancer *Category* = Illness **Episodic**	*N* = 250 *Geo* = US *Pop = N/A*	*IR* = keyword sample *Type* = AGS *Coding* = CA *Analysis* = O R χ^2	**N/A**
Niederdeppe [244]	*Topic* = Nutrition (transfats) *Category* = Policies&Inst **Non-episodic**	*N* = 361 *Geo* = US *Pop = N/A*	*IR* = keyword *Type* = AGS *Coding* = CA *Analysis* = TS R	**N/A**
Niederkrotenthaler [104]	*Topic* = Suicide *Category* = Injury **Non-episodic**	*N* = 2016 *Geo* = Austria *Pop = N/A*	*IR* = keyword *Type* = INF *Coding* = CA *Analysis* = O R χ^2	**-**
Niederkrotenthaler [94]	*Topic* = Suicide *Category* = Injury **Non-episodic**	*N* = 179 *Geo* = Austria *Pop = N/A*	*IR* = keyword manual *Type* = INF *Coding* = CA *Analysis* = O R χ^2	**N/A**

Author	Topic Research Category Episodic news coverage	Corpus Size Geography Population	Methods	News Quality
O'Hara[97]	*Topic* = Eating disorders *Category* = Illness **Non-episodic**	*N* = 252 *Geo* = US *Pop = N/A*	*IR* = keyword *Type* = AGS INF FR *Coding* = CA *Analysis = #*	**-**
Payne[96]	*Topic* = Elder/child abuse (crime) *Category* = Injury **Non-episodic**	*N* = 11630 *Geo* = US *Pop* = Elderly, children	*IR* = keyword *Type* = AGS FR *Coding* = CA *Analysis = #*	**-**
Pirkis[105]	*Topic* = Suicide *Category* = Injury **Non-episodic**	*N* = 80 *Geo* = Australia *Pop = N/A*	*IR* = manual *Type* = AGS INF *Coding* = keyword *Analysis = #*	**-**
Pirkis[60]	*Topic* = Suicide *Category* = Injury **Non-episodic**	*N* = 5109 *Geo* = Australia *Pop = N/A*	*IR* = keyword sample *Type* = INF *Coding* = CA *Analysis* = χ^2	**+**
Quick[93]	*Topic* = Steroid use in sports *Category* = Risk **Non-episodic**	*N* = 1062 *Geo* = US *Pop = N/A*	*IR* = keyword sample *Type* = INF *Coding* = CA *Analysis* = Q M χ^2	**-**
Rabi [245]	*Topic* = Diabetes medication (rosiglitazone) *Category* = Treatment **Episodic**	*N* = 156 *Geo* = Worldwide * *Pop = N/A*	*IR* = keyword manual *Type* = INF FR *Coding* = CA *Analysis* = χ^2	**+/-**
Racine[246]	*Topic* = End of life decision making *Category = Psych* **Episodic**	*N* = 1141 *Geo* = US *Pop = N/A*	*IR* = keyword manual *Type* = INF FR *Coding* = CA *Analysis = #*	**-**

* English language news

Author	Topic Research Category Episodic news coverage	Corpus Size Geography Population	Methods	News Quality
Racine [90]	*Topic* = Medical science (neuroscience) *Category* = Research **Non-episodic**	*N* = 1256 *Geo* = US & UK *Pop = N/A*	*IR* = keyword *Type* = AGS INF FR *Coding* = CA *Analysis* = F	**N/A**
Rada[247]	*Topic* = Medical research retractions *Category* = Research **Non-episodic**	*N* = 422 *Geo = N/A* *Pop = N/A*	*IR* = keyword manual *Type* = AGS *Coding* = keyword *Analysis = #*	**N/A**
Raman[248]	*Topic* = Medical ethics *Category* = Research **Non-episodic**	*N* = 14 *Geo* = India *Pop = N/A*	*IR* = keyword *Type* = FR *Coding* = CA *Analysis = #*	**-**
Reid[249]	*Topic* = Cosmetic Surgury *Category* = Treatment **Non-episodic**	*N* = 1191 *Geo* = UK *Pop = N/A*	*IR* = keyword *Type* = AGS FR *Coding* = CA *Analysis = #*	**+/-**
Rosales[250]	*Topic* = Motor vehicle crashes *Category* = Injury **Non-episodic**	*N* = 473 *Geo* = US *Pop = N/A*	*IR* = sample *Type* = FR INF *Coding* = CA *Analysis = #*	**-**
Schwartz[107]	*Topic* = Illegal Drug Use (Methamphetamine) & Sexual health *Category* = Risk **Non-episodic**	*N* = 168 *Geo* = US *Pop* = MSM	*IR* = keyword sample *Type* = FR *Coding* = CA *Analysis* = χ^2	**-**
Schwitzer[59]	*Topic* = Medical interventions *Category* = Treatment **Non-episodic**	*N* = 500 *Geo* = US *Pop = N/A*	*IR* = manual *Type* = INF *Coding* = CA *Analysis = #*	**-**

Author	Topic Research Category Episodic news coverage	Corpus Size Geography Population	Methods	News Quality
Scully[92]	*Topic* = Cancer (Melanoma) *Category* = Prevention **Non-episodic**	*N* = 547 *Geo* = Australia *Pop = N/A*	*IR* = keyword manual *Type* = FR *Coding* = CA *Analysis* = R	**+/-**
Seale[251]	*Topic* = Sleep disorders *Category* = Illness **Non-episodic**	*N* = 1051 *Geo* = UK *Pop = N/A*	*IR* = keyword *Type* = FR *Coding* = auto *Analysis* = #*	**-**
Slater[87]	*Topic* = Cancer *Category* = Illness **Non-episodic**	*N* = 274 *Geo* = US *Pop = N/A*	*IR* = manual sample *Type* = AGS FR *Coding* = CA *Analysis* = Z	**+/-**
Slopen[252]	*Topic* = Mental Illness *Category* = Illness **Non-episodic**	*N* = 1252 *Geo* = US *Pop* = Children	*IR* = keyword *Type* = INF FR *Coding* = CA *Analysis* = χ^2	**+/-**
Smith[253]	*Topic* = Smoking *Category* = Risk **Non-episodic**	*N* = 8390 *Geo* = US *Pop* = Adolescents	*IR* = keyword *Type* = AGS FR *Coding* = CA *Analysis* = O χ^2	**N/A**
Smith[78]	*Topic* = MMR Vaccine *Category* = Prevention **Episodic**	*N* = 1700 † *Geo* = US *Pop = N/A*	*IR* = keyword *Type* = AGS *Coding* = CA *Analysis* = O R	**N/A**

* Comparative keyword analysis
† Specific number not stated in research article, this is an estimate based on a graph in the article.

Author	Topic Research Category Episodic news coverage	Corpus Size Geography Population	Methods	News Quality
Stephenson [254]	*Topic* = Pandemic influenza *Category* = Illness **Non-episodic**	*N* = 333 *Geo* = Australia *Pop = N/A*	*IR* = keyword manual sample *Type* = INF FR *Coding* = CA *Analysis* = #	**N/A**
Stryker[88]	*Topic* = Cancer control *Category* = Illness **Non-episodic**	*N* = 5892 *Geo* = US *Pop* = Various ethnicities [*]	*IR* = keyword manual *Type* = INF AGS *Coding* = CA *Analysis* = t χ^2	**+/-**
Stryker[255]	*Topic* = Cancer *Category* = Prevention **Non-episodic**	*N* = 954 *Geo* = US *Pop = N/A*	*IR* = keyword sample *Type* = AGS *Coding* = CA *Analysis* = O R	**N/A**
Stryker [197]	*Topic* = Cancer (ethnic) *Category* = Risk **Non-episodic**	*N* = 5892 *Geo* = US *Pop* = Various ethnicities[†]	*IR* = keyword manual sample *Type* = INF *Coding* = CA *Analysis* = O χ^2	**+/-**
Takahashi[256]	*Topic* = Asbestos *Category* = Risk **Episodic**	*N* = 329 [‡] *Geo* = Japan *Pop = N/A*	*IR* = keyword *Type* = AGS FR *Coding* = CA *Analysis* = #[§]	**N/A**
Taylor-Clark[29]	*Topic* = Health Disparities *Category* = Disparities **Non-episodic**	*N* = 194 *Geo* = US *Pop* = Blacks	*IR* = keyword *Type* = AGS FR *Coding* = CA *Analysis* = #	**+**

[*] Ethnicities represented in Lexis-Nexis' Ethnic NewsWatch database.
[†] Ethnicities represented in Lexis-Nexis' Ethnic NewsWatch database.
[‡] Just headlines, not full news article text.
[§] Network analysis of word cooccurence

Author	Topic Research Category Episodic news coverage	Corpus Size Geography Population	Methods	News Quality
Tong[257]	*Topic* = Chronic Kidney Disease *Category* = Illness **Non-episodic**	*N* = 214 *Geo* = Australia *Pop = N/A*	*IR* = keyword manual *Type* = FR *Coding* = CA *Analysis* = #	**-**
Tonmyr [82]	*Topic* = Child maltreatment *Category* = Injury **Episodic**	*N* = 29 *Geo* = Canada *Pop* = Children	*IR* = keyword *Type* = INF FR *Coding* = CA *Analysis* = #	**N/A**
Weidner[108]	*Topic* = Drug abuse (Methamphetamine) *Category* = Risk **Non-episodic**	*N* = 1150 *Geo* = US *Pop = N/A*	*IR* = keyword *Type* = AGS INF *Coding* = CA *Analysis* = #	**+/-**
Wilson[258]	*Topic* = Cancer medication, (Trastuzumab, Herceptin) *Category* = Treatment **Episodic**	*N* = 361 *Geo* = UK *Pop = N/A*	*IR* = manual *Type* = AGS INF FR *Coding* = CA *Analysis* = #	**-**
Wilson [89]	*Topic* = Medical interventions *Category* = Treatment **Non-episodic**	*N* = 1230 *Geo* = Australia *Pop = N/A*	*IR* = keyword manual pre *Type* = INF FR *Coding* = CA *Analysis* = R A	**+/-**
Yong [259]	*Topic* = Medication black-box warnings *Category* = Policies&Inst **Non-episodic**	*N* = 551 *Geo* = US *Pop = N/A*	*IR* = keyword *Type* = INF FR *Coding* = CA *Analysis* = F R U	**-**

APPENDIX B. WEB INTERFACE USED TO COLLECT SUBJECT RESPONSES

B.1 Study Website Overview

All coder responses for this study were collected using a content analysis website specifically constructed for this study. The site was located online at the domain www.salientnews.com. This software is the basis for the SalientHealthNews User Annotation/Feedback module.

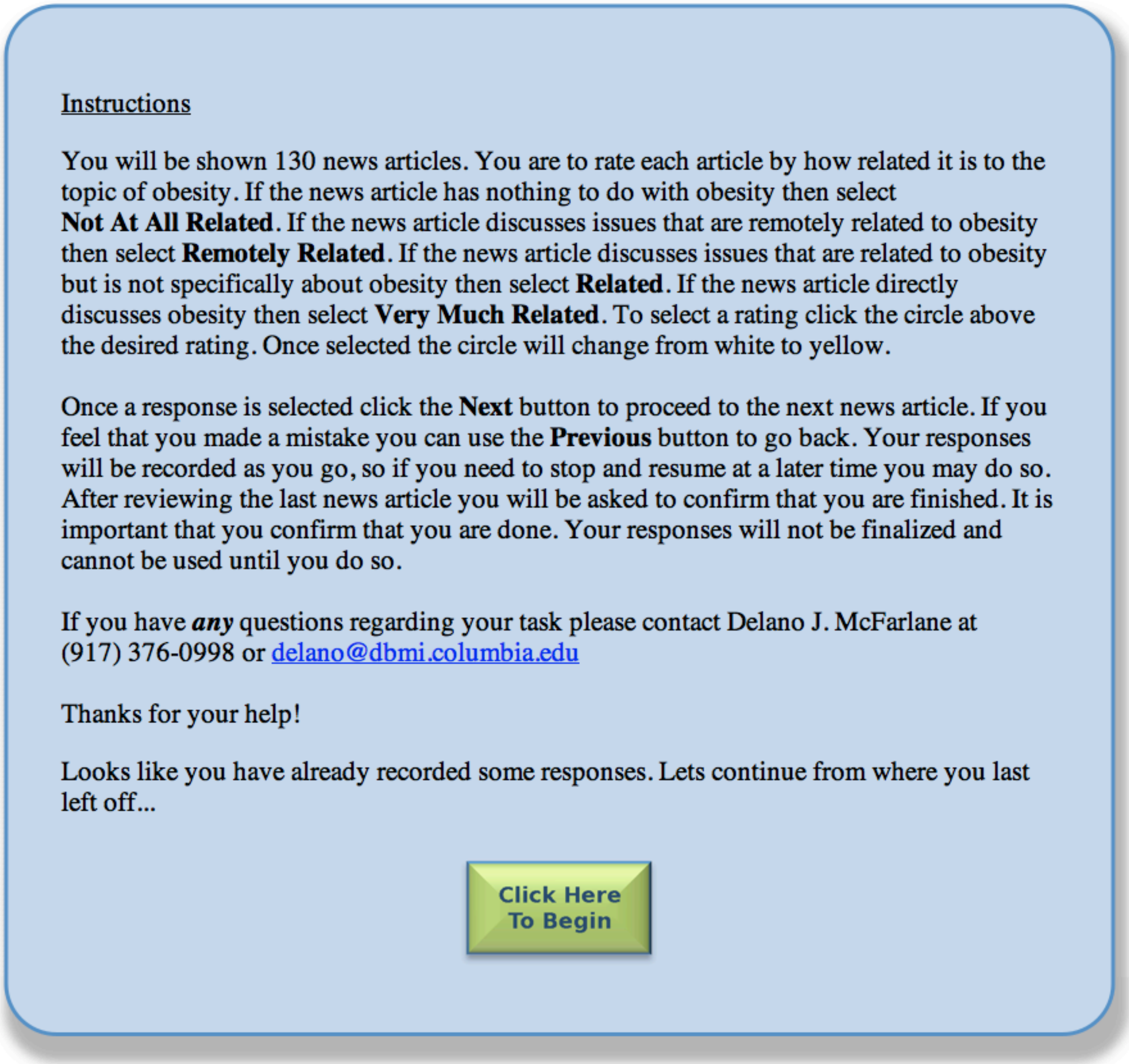
Preceding each screen shot are the instructions that were displayed to participants. Once someone agreed to participate in the study they were sent an email that contains an Internet URL that took them to either the framing or obesity relevance coding task.

B.2 Obesity News Classification Task

Instructions:

> *"You will be presented with a series of news articles. For each news article please indicate how much that news article is related to the topic of obesity. You will have 4 options to choose from when selecting your response; (1) not at all related to obesity, (2) remotely related to obesity, (3) somewhat related to obesity, (4) very much related to obesity. Once a response is recorded you may click the **Next** button to proceed to the next news article. If you feel that you made a mistake you can use the **Previous** button to go back. Your responses will be recorded as you go, so if you need to stop and resume at a later time you may do so. After reviewing the last news article you will be asked to confirm that you are finished. It is important that you confirm that you are done. Your responses will not be finalized and cannot be used until you do so. Thanks for your help!"*

Obesity news coding task: instructions



Instructions

You will be shown 130 news articles. You are to rate each article by how related it is to the topic of obesity. If the news article has nothing to do with obesity then select **Not At All Related**. If the news article discusses issues that are remotely related to obesity then select **Remotely Related**. If the news article discusses issues that are related to obesity but is not specifically about obesity then select **Related**. If the news article directly discusses obesity then select **Very Much Related**. To select a rating click the circle above the desired rating. Once selected the circle will change from white to yellow.

Once a response is selected click the **Next** button to proceed to the next news article. If you feel that you made a mistake you can use the **Previous** button to go back. Your responses will be recorded as you go, so if you need to stop and resume at a later time you may do so. After reviewing the last news article you will be asked to confirm that you are finished. It is important that you confirm that you are done. Your responses will not be finalized and cannot be used until you do so.

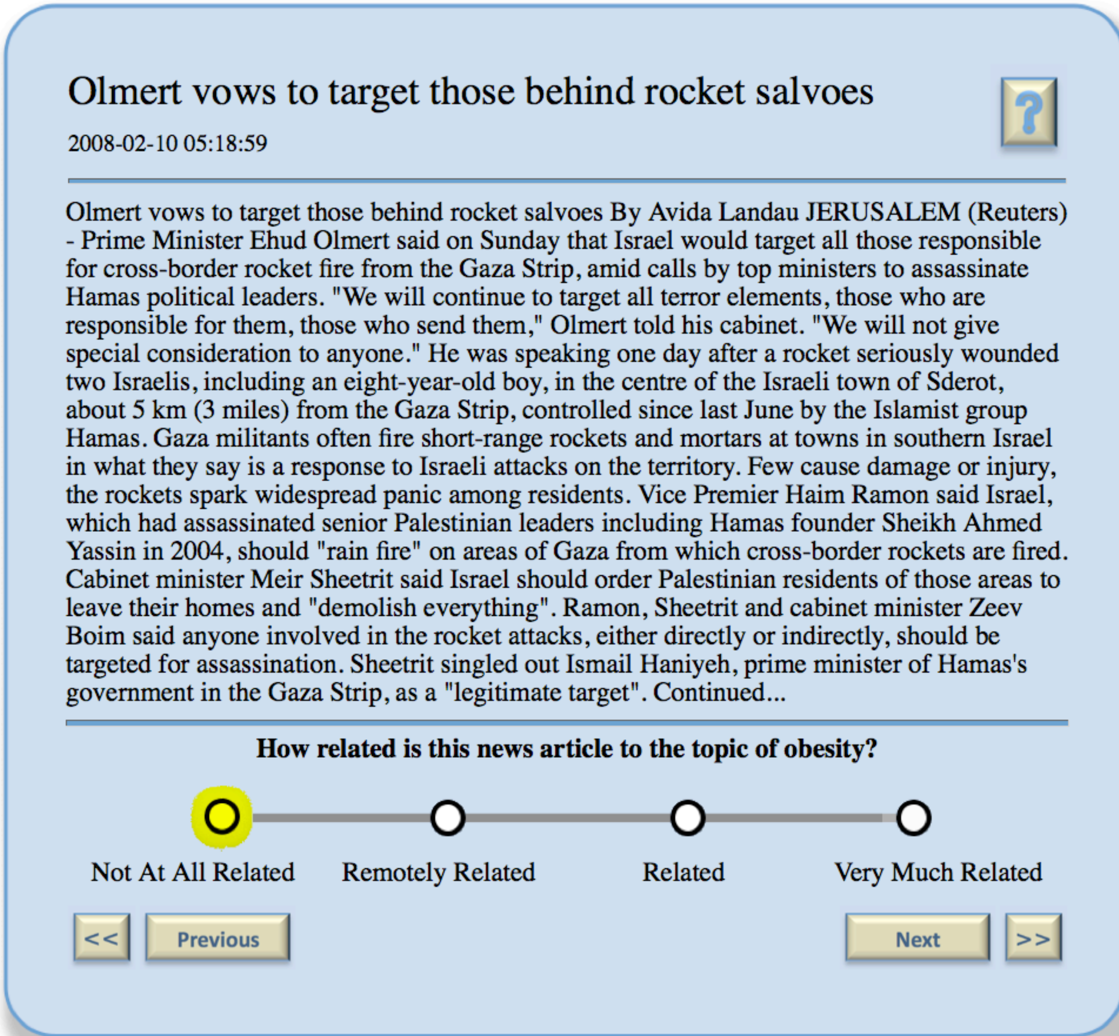
If you have ***any*** questions regarding your task please contact Delano J. McFarlane at (917) 376-0998 or delano@dbmi.columbia.edu

Thanks for your help!

Looks like you have already recorded some responses. Lets continue from where you last left off...

Click Here
To Begin

Obesity news coding task: coding



Olmert vows to target those behind rocket salvoes

2008-02-10 05:18:59

Olmert vows to target those behind rocket salvoes By Avida Landau JERUSALEM (Reuters) - Prime Minister Ehud Olmert said on Sunday that Israel would target all those responsible for cross-border rocket fire from the Gaza Strip, amid calls by top ministers to assassinate Hamas political leaders. "We will continue to target all terror elements, those who are responsible for them, those who send them," Olmert told his cabinet. "We will not give special consideration to anyone." He was speaking one day after a rocket seriously wounded two Israelis, including an eight-year-old boy, in the centre of the Israeli town of Sderot, about 5 km (3 miles) from the Gaza Strip, controlled since last June by the Islamist group Hamas. Gaza militants often fire short-range rockets and mortars at towns in southern Israel in what they say is a response to Israeli attacks on the territory. Few cause damage or injury, the rockets spark widespread panic among residents. Vice Premier Haim Ramon said Israel, which had assassinated senior Palestinian leaders including Hamas founder Sheikh Ahmed Yassin in 2004, should "rain fire" on areas of Gaza from which cross-border rockets are fired. Cabinet minister Meir Sheetrit said Israel should order Palestinian residents of those areas to leave their homes and "demolish everything". Ramon, Sheetrit and cabinet minister Zeev Boim said anyone involved in the rocket attacks, either directly or indirectly, should be targeted for assassination. Sheetrit singled out Ismail Haniyeh, prime minister of Hamas's government in the Gaza Strip, as a "legitimate target". Continued...

How related is this news article to the topic of obesity?

B.3 News frame coding task

Instructions for first round of coding:

*"You will be presented with a series of articles about obesity. You will also be given 8 health news categories. For each article select all categories that match. If none of the categories match the article then select **NO MATCH**. After you have entered a response for a news article you can continue to the next article by pressing the **Next** button. If you want to edit an earlier response use the **Previous** button. After entering a response for all news articles press the **Done** button and then confirm that you are finished. It is important that you confirm that you are done. Your responses will not be finalized and cannot be used until you do so. Your responses will be saved as you work so if you need to finish later you can pick up where you left off. However you are encouraged to finish in as few sessions as possible. Thanks for your help!"*

News framing coding task: instructions

Instructions

You will be shown 80 health news articles related to the topic of obesity. You will also be given 8 categories (listed below). For each news article select all categories that apply.

Category	Description
Personal Responsibility	News articles about an individual's personal role and responsibility in maintaining their health
Environmental Risk	News articles about the risks imposed upon a person or community by their environment (where they live, work, etc...)
Economic Issues	News articles about potential or realized economic issues due to exposure, treatment or health consequences
Genetic/Biological Issues	News articles about the biological or genetic (hereditary) causes or contributors to health
Psychosocial Treatments	News articles about the impact of treatments that try to modify the psychological and social aspects of an individual's behavior
Medical Treatments	News articles about the impact of new and existing medical treatments (e.g. surgury, drugs)
Lifestyle	News articles about the role an individual's activities, attitudes and values plays in their health
Culture / Community	News articles about the role that the attitudes, beliefs and behaviors of an individual's family, friends and associates plays in that person's health

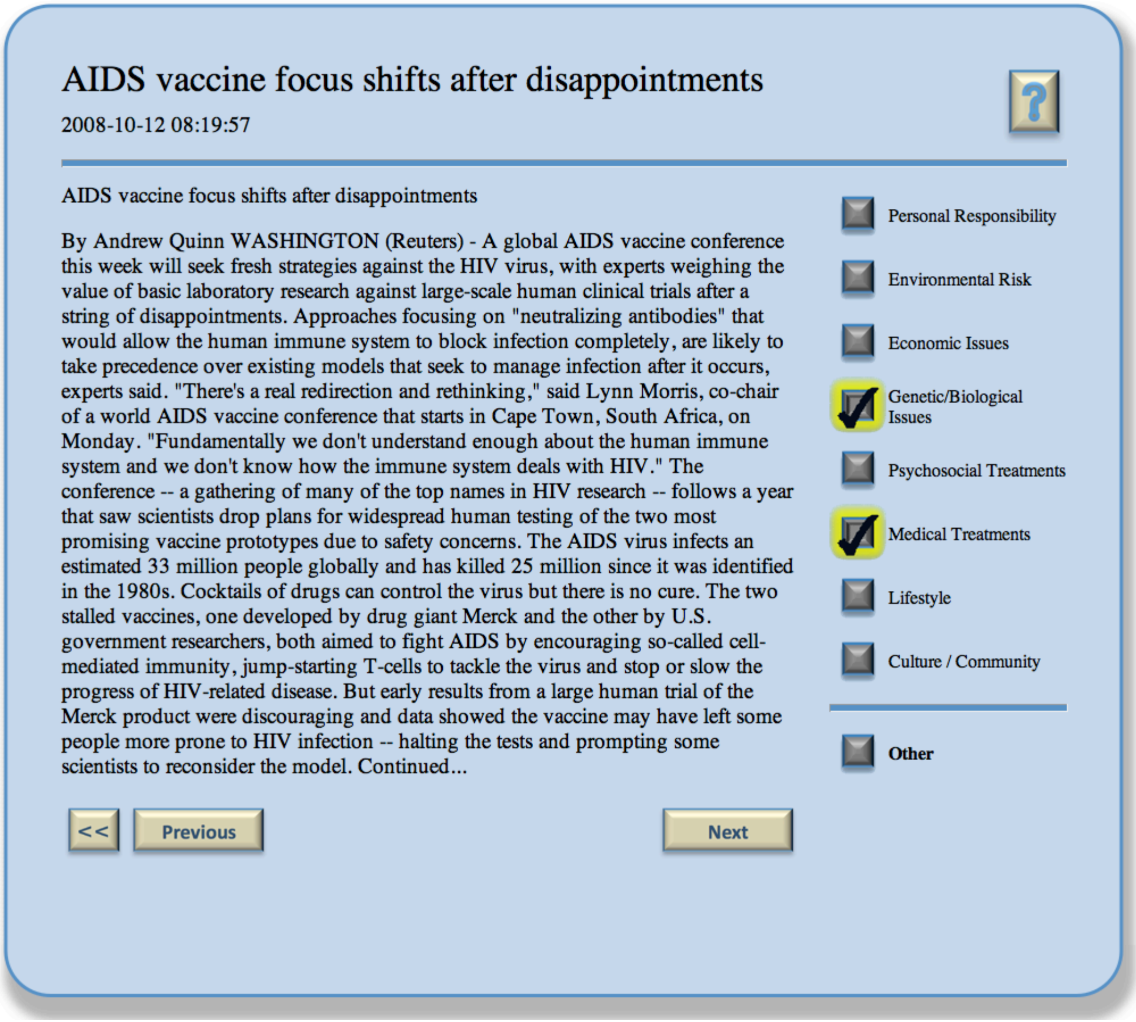
If none of these categories apply then select **Other** and type a short description in the area that is provided (only a couple words are needed). To be reminded of a category's description place your mouse over the category's checkbox.

After you have entered a response for a news article you can continue to the next article by pressing the **Next** button. If you want to edit an earlier response use the **Previous** button. After responding to all news articles press the **Done** button and then confirm that you are finished. It is important that you confirm that you are done. Your responses will not be finalized and cannot be used until you do so. Your responses will be saved as you work so if you need to finish later you can pick up where you left off. However you are encouraged to finish in as few sessions as possible. Thanks for your help!

Looks like you have already recorded some responses. Lets continue from where you last left off...

Click Here
To Begin

News framing coding task: coding



Framing coding task (2nd round): instructions

Instructions

The purpose of this exercise is to increase the level of agreement for the responses previously provided by you and others, and to gain information on possible reasons behind disagreements between your responses and those of the other participants. While the amount of agreement for the previous responses was good, higher levels of agreement and an understanding of the causes behind disagreements would greatly benefit the research being conducted.

In the following exercise you will be shown health news articles that you previously stated were related or not related to a set of categories. For each article at least one respondant gave a response different from yours for the two categories listed below. You will now be asked to focus just on these two categories and again give your opinion on if news articles are related or not related to them.

Category	Description
Genetic/Biological Issues	News articles about the biological or genetic (hereditary) causes or contributors to health
Lifestyle	News articles about the role an individual's activities, attitudes and values plays in their health

Along with each news article you will also be shown your previous response. Your current responses should be based on the news article itself, the category definitions, your past response and the knowledge that others disagreed with you. Keep in mind that your previous response may have been in the majority, so feel free to re-enter your previous response if you still feel that it is correct.

You will also be given an area where you can enter a comment. Comments are optional but may help in understanding possible causes of disagreement between respondants.

Once you have entered your response for a news article, scroll to the bottom of the page and press the **Next** button to continue to the next news article. If you want to go back and edit an earlier response use the **Previous** button. After responding to all news articles press the **Done** button and then confirm that you are finished. It is important that you confirm that you are done. Your responses will not be finalized and cannot be used until you do so. Your responses will be saved as you work so if you need to finish later you can pick up where you left off.

If you have ***any*** questions regarding the use of this website or the task that you have been asked to perform, ***please*** contact me via email at [redacted] or telephone at [redacted]. Thanks for your help!

Looks like you have already recorded some responses. Lets continue from where you last left off...

Click Here To Begin

Framing coding task (2nd round): coding

Weight loss may not harm obese teens' bones

2008-02-14 14:39:42

Although adults who lose weight may also lose some bone mass, obese adolescents seem to keep gaining bone density as they shed pounds, a study suggests.

The findings, reported in the medical journal Obesity, offer some reassurance that obese children"s weight loss may not come at the expense of their bone health.

The study included 62 obese adolescents who completed an intensive year-long weight-loss program and had X-rays to chart changes in their bone mineral content.

The researchers found that even as the teenagers lost weight, their bone mass continued to increase and remained higher than that of a comparison group of thin adolescents.

"Our findings suggest that successful, medically supervised obesity treatment in adolescents does not cause major problem for bone health," said lead researcher Dr. Nicolas Stettler of The Children"s Hospital of Philadelphia.

Given the many health benefits of weight loss for obese children, concerns about bone development should not keep them from getting help for their weight, Stettler told Reuters Health.

Some past studies have found that overweight and obese children are at greater risk of bone fractures than their normal-weight peers.

The reasons are not certain, since a few studies, including the current one, show that obese adolescents typically have greater bone mass than thinner teens do.

Factors unrelated to bone mass may help explain the higher fracture risk, Stettler and his colleagues note.

For example, falls may be harder on obese children owing to their extra weight or poorer coordination.

In theory, losing weight might lower obese kids" odds of sustaining a bone fracture, according to Stettler.

But, he said, this has not yet been proven in studies.

<< Previous Next

Last time you said this news article

DOES NOT DISCUSS Genetic/Biological Issues
- and -
DOES NOT DISCUSS Lifestyle

At least one expert disagrees with you.

What is your opinion of this news article now?

This news article...

DISCUSSES Genetic/Biological Issues

Does not discuss Genetic/Biological Issues

-

DISCUSSES Lifestyle

Does not discuss Lifestyle

Please enter a few words that describe this article:

This was a difficult one to code, but I think I got it right this time.

429 chars left

APPENDIX C. EXAMPLES FROM THE DAL AND LIWC DICTIONARIES

This appendix contains examples from the DAL and LIWC dictionaries.

C.1 Dictionary of Affect in Language (DAL) examples

Every word in the dictionary is scored on a 3-point scale for each of the 3 dimensions. For *Activation* the ratings mean 1) passive, 2) in between, and 3) active. For *Imagery* the ratings mean 1) difficult to envision, 2) in between, and 3) easy to envision. For *Pleasantness* the ratings mean 1) unpleasant, 2) in between, and 3) pleasant. Below are 50 randomly selected words from the DAL dictionary along with their scores:

DAL examples

Word	Pleasantness	Activation	Imagery
apologizing	2.00	2.11	1.60
became	2.00	1.75	1.20
churches	2.00	1.43	3.00
competition	1.88	2.56	3.00
confidence	2.80	2.27	1.80
daren*t	1.38	1.73	1.20
defined	1.75	1.67	1.20
didn*t	1.78	1.60	1.00
direct	1.88	1.44	1.00
drying	1.82	1.91	2.60
effigy	1.73	1.55	1.00
fish	1.67	1.86	3.00
herself	2.25	1.50	1.80
horses	2.44	2.00	3.00
interest	2.67	1.25	1.60
intervention	1.50	1.75	1.60
interviews	2.00	1.92	2.20
kiss	2.80	2.33	3.00
language	1.90	1.86	1.20
lowered	1.86	1.70	2.00
minerals	1.67	1.67	2.40
mr	1.71	1.57	1.20
one-third	1.36	1.45	2.20
particles	1.44	1.89	1.60
points	1.43	1.29	2.20
populated	1.67	2.22	1.80

DAL examples

Word	Pleasantness	Activation	Imagery
presidents	1.80	1.90	1.80
proceed	1.78	2.00	1.20
pupils	1.71	2.00	2.20
ratty	1.50	2.00	1.60
reader	2.00	1.83	2.00
regretted	1.00	1.71	1.20
republic	1.67	1.44	1.40
rhythm	2.44	1.78	2.00
rod	1.89	1.67	2.40
satisfactorily	2.57	1.17	1.40
seems	1.71	1.50	1.00
select	1.78	2.00	1.80
shift	1.17	2.00	2.00
specialists	2.29	2.33	1.60
statute	1.71	1.67	1.00
strategic	2.38	2.00	1.80
to	1.78	1.57	1.20
tones	2.00	1.63	1.40
transportation	2.17	2.40	2.80
trial	1.25	2.14	1.80
unfairness	1.50	1.67	1.20
very	1.83	1.80	1.00
visited	2.78	2.00	1.40
wood	2.13	1.43	2.80

C.2 Linguistic Frequency and Word Count (LIWC) examples

Major categories (e.g. *Negative emotions*) are comprised of all terms in their subcategories (for *Negative emotions* that would be all terms in the subcategories *anger, anxiety*, and *sad*).

LIWC examples

categories and subcategories	***Examples*** *('*' = wildcard)*
Social Processes	*blam*, bro, cells, citizens, help, his, marriag*, motherly, shares*
Family	*exhusband*, exwive*, families*, grandchil*, ma, mothers, niece*, pappy, son's, stepfat**
Friend	*bud, buddies*, companion, comrad*, fellow*, friend*, girlfriend*, mate's, partner*, sweetie**
Humans	*child, gentlem*, girls*, grownup*, male, participant*, person's, self, woman's, women**
Affective processes	*aversi*, harming, honest*, invigor*, sufferer*, superior*, tense*, thrill*, uneas*, unsavo**
Negative emotions	*contempt*, disturb*, fuming, hopeless*, insecur*, reject*, sadde*, scary, shock*, trauma**
Anger	*abuse*, annoy*, danger*, disgust*, fiery, frustrat*, maniac*, rebel*, spite*, wicked**
Anxiety	*apprehens*, distress*, inhib*, nervous*, overwhelm*, panic*, repress*, terror*, uncontrol*, upset**
Sad	*alone, cried, grief, griev*, inadequa*, lost, piti*, suffering, weep*, yearn**
Positive emotions	*agree, careful*, deligh*, helping, laidback, prize*, support, toleran*, worthwhile, yay*
Cognative processes	*add, bet, cohere*, comply*, couldnt, indefinit*, neglect*, reason*, solved, sure**
Causation	*because, control*, create*, depended, depends, led, makes, origin, purpose*, used*
Certainty	*absolute, apparent, clearly, definitely, everything*, fact, facts, must, obvious*, truly*
Discrepancy	*desir*, hopefully, ideal*, if, mustnt, need, oughtn't, regret*, would, wouldve*
Exclusive	*but, except, just, not, or, rather, really, something*, vs, whether*
Inclusive	*addit*, along, and, around, close, come, each, into, we, with*
Inhibition	*abandon*, banning, bound*, conflict*, constrict*, rein*, restrain*, retain*, stop, wariness*
Insight	*believe, decis*, evaluat*, find*, grasp*, infers, knows, lesson*, think, thinks*
Tentative	*anybod*, guess, guesses, luck, might've, most, perhaps, sort, supposes, unresolv**
Perception	*edges, flavor*, inaudibl*, limp*, listened, oil*, ring, rubbed, sugar*, yelling*

LIWC examples

categories and subcategories	***Examples*** *('*' = wildcard)*
Feel	*brush*, cool, drie*, freez*, gripp*, heavy*, loose*, presses, sands, weighted*
Hear	*ear, hear, hearing, listening, noise, quiet*, sang, sound*, spoke*, yelling*
See	*beaut*, color*, pink*, round*, saw, see, seeing, seer, stare*, view*
Biological processes	*abortion*, bald, bladder*, cough*, dosage, eaten, helpings, lung*, sicker, wart*
Body	*brain*, drool*, facial*, finger*, jaw*, joints, neural*, scalp, sweat*, thirst**
Health	*burp*, chills, clinic*, colono*, glaucoma, gynecolog*, hemor*, paining, poison*, vitamin**
Ingestion	*binge*, brunch*, candie*, cigar*, dined, dinner*, eats, grocer*, lunch*, whisky**
Sexual	*abortion*, aids, genital*, hug, kiss*, lover*, lust*, nude*, penis*, prudish**
Personal concerns	
Home	*bedroom*, domestic*, family, garden*, home, mortg*, neighbor*, pet, residen*, roomate**
Leisure	*entertain*, hangout, karaoka, margarita*, restau*, skiing, tv*, vacation*, weekend*, wine*
Work	*business*, commuting, duty, employ*, financ*, mgmt, outsourc*, overpaid, overworked, pay**

APPENDIX D. NEWS ARTICLE EXAMPLES

Title: **Exercise combats cancer-related fatigue: report**

Publish Date: **2008-04-22**

http://feeds.reuters.com/~r/reuters/healthNews/~3/275526964/idUSCOL26012020080422

Nearly all cancer patients experience fatigue, Dr. Fiona Cramp and colleagues note in the latest issue of The Cochrane Library, a publication of The Cochrane Collaboration, an international organization that evaluates medical research.

According to guidelines from the National Comprehensive Cancer Network, treatable factors that may be related to cancer-related fatigue, such as pain, emotional distress, sleep disturbance, anemia, nutrition, activity level, and co-morbid illnesses, should be identified and treated.

However, there is no consensus regarding the effect of exercise on cancer-related fatigue once treatable causes have been addressed.

Cramp, of the University of the West of England in Bristol, UK, and colleagues searched the medical literature for controlled trials that evaluated the effect of exercise on cancer-related fatigue. They identified 28 studies involving 2083 participants. More than half of the studies involved women with breast cancer.

"Statistically significant improvements in fatigue were identified following an exercise program carried out either during cancer therapy or following cancer therapy," the researchers report. Most programs involved moderate-intensity exercise performed two or three times per week.

Cramp''s team recommends that exercise be considered as one of several components of the management strategy for cancer-related fatigue, which may also include other nonpharmacologic interventions, including psychological and social therapies, stress management, nutrition therapy and sleep therapy.

"Exercise shouldn't be used in isolation but should definitely be included as one of the components in the package of interventions used during and after treatment," Cramp said in a written statement.

Title: **Nonfat milk linked to prostate cancer**

Publish Date: **2008-01-02**

http://feeds.reuters.com/~r/reuters/healthNews/~3/210044591/idUSHAR26781420080102

The amount of calcium and vitamin D in the diet appears to have little or no impact on the risk of prostate cancer, but the consumption of low-fat or nonfat milk may increase the risk of the malignancy, according to the results of two studies published in the American Journal of Epidemiology.

Dietary calcium and dairy products have been thought to increase the risk of prostate cancer by affecting vitamin D metabolism. Data from several prospective studies have supported an association, but many other studies have failed to establish a link.

To explore this topic further, Dr. Song-Yi Park, from the University of Hawaii in Honolulu, and colleagues, analyzed data from subjects enrolled in the Multiethnic Cohort Study. This study, conducted between 1993 and 2002, included adults between 45 and 75 years old, were primarily from five different ethnic or racial groups, and lived in California or Hawaii.

A total of 82,483 men from the study completed a quantitative food frequency questionnaire and various factors, such as weight, smoking status, and education levels were also noted, Park's group said.

During an average follow-up period of 8 years, 4,404 men developed prostate cancer. There was no evidence that calcium or vitamin D from any source increased the risk of prostate cancer. This held true across all racial and ethnic groups.

In an overall analysis of food groups, the consumption of dairy products and milk were not associated with prostate cancer risk, the authors found. Further analysis, however, suggested that low-fat or nonfat milk did increase the risk of localized tumors or non-aggressive tumors, while whole milk decreased this risk.

In a similar analysis, Dr. Yikyung Park, from the National Cancer Institute at National Institutes (NIH) of Health in Bethesda, Maryland, and colleagues investigated the relationship of calcium and vitamin D and prostate cancer in 293,888 men enrolled in the NIH-American Association of Retired Persons Diet and Health Study, conducted between 1995 and 2001. The average follow-up period was 6 years.

No link between total or supplemental dietary calcium and the total number of non-advanced prostate cancer cases was noted. Total calcium intake was tied to advanced and fatal disease, but both associations fell short of statistical significance.

Similar to the first study's findings, skim milk was linked with advanced prostate cancer. Calcium from non-dairy food, by contrast, was tied to a reduced risk of non-advanced prostate cancer.

"Our findings do not provide strong support for the hypothesis that calcium and dairy foods increase the risk of prostate cancer. The results from other large...studies, with adequate numbers of advanced and fatal prostate cancers, may shed further light on this question," Park's team concludes.

SOURCE: American Journal of Epidemiology, December 1, 2008.

Title: **High fat diet may abet prostate cancer progression**

Publish Date: **2008-07-02**

http://feeds.reuters.com/~r/reuters/healthNews/~3/325097396/idUSCOL26344120080702

Diets high in saturated fat may increase the risk of prostate cancer progression, researchers from the University of Texas M. D. Anderson Cancer Center in Houston report.

In a follow up study of men who had their cancerous prostates removed, researchers found that men who consumed higher amounts of saturated fat -- mostly from steaks, burgers, cheese, ice cream, salad dressings, and mayonnaise -- were nearly two times more likely to experience disease progression after surgery than men with lower saturated fat intake.

"Diet before surgery, especially saturated fat, may modulate patient outcome after surgery," Dr. Sara S. Strom, who was involved in the study, told Reuters Health.

Strom and colleagues also found significantly shorter "disease-free" survival times among obese men who ate high amounts of saturated fat compared with non-obese men consuming diets low in saturated fat.

These results expand upon the team's previous finding linking obesity with prostate cancer progression "and suggest that saturated fat intake plays a role in prostate cancer progression," the researchers note in the International Journal of Cancer.

Strom's group used standard food questionnaires to assess the saturated fat intake of 390 men during the year before surgery for localized, or "organ-confined" prostate cancer. The researchers also assessed the men's medical and family history for other risk factors for disease progression.

The men, all Caucasian, were about 60 years old on average and consumed between 600 and 5,000 calories daily. Overall, 293 men averaged 10 percent of their daily energy from saturated fat (low intake) while 97 men averaged 14 percent (high intake).

Obese men with a high saturated fat intake had the shortest survival time free of prostate cancer (19 months), while non-obese men with low intake survived the longest time free of the disease (46 months).

Non-obese men with high intake and obese men with low intake had "disease-free" survival of 29 and 42 months, respectively, the researchers report.

Additional investigations looking at associations between post-surgery dietary changes and disease progression would be worthwhile, Strom suggests.

SOURCE: International Journal of Cancer, June 1, 2008

Title: **Birth control pills may lower colon cancer risk**

Publish Date: **2008-02-20**

http://feeds.reuters.com/~r/reuters/healthNews/~3/238257429/idUSKIM05941720080220

Women who have used birth control pills seem to have a slightly decreased risk of colon cancer as they age, a new study suggests.

Researchers found that among nearly 90,000 women ages 40 to 59, those who had ever used oral contraceptives were 17 percent less likely to develop colon cancer over the next 16 years.

The findings, which appear in the International Journal of Cancer, are in line with evidence suggesting that estrogen plays a role in colon cancer risk.

Some studies, for example, have found that older women on hormone replacement therapy (HRT) have a lower risk of the disease. In addition, lab experiments have shown that estrogen may inhibit tumor development in the colon by affecting cell growth, or by lowering levels of a cancer-linked hormone called IGF-1.

However, it's too soon to conclude that birth control pills offer colon cancer protection, according to lead researcher Dr. Geoffrey C. Kabat, of Albert Einstein College of Medicine in New York.

For one, he told Reuters Health, the risk reduction was small. In addition, the study found no "dose-response" relationship between oral contraceptives and colon cancer -- that is, the risk reduction was not greater among women who'd used birth control pills for longer periods. In general, a positive dose-response strengthens the case for cause-and-effect relationships.

It's also possible that there is something else about women who use birth control pills that makes them less susceptible to colon cancer, Kabat explained. They may, for example, be more physically active and weigh less -- two factors that studies suggest may lower the risk of colon cancer.

Even though the researchers attempted to account for lifestyle habits and other factors in their analysis, Kabat said they cannot exclude those things as an explanation for their findings.

The results are based on 89,835 Canadian women taking part in a study on breast cancer screening that followed them for an average of 16 years. During that time, women who had ever used birth control pills were less likely to develop colon cancer. However, this was not true of women who had used HRT, in contrast to what several previous studies have found.

According to Kabat, there are still "many questions to sort out" regarding hormone use and colon cancer, and the results of any single study have to be interpreted cautiously. More studies are needed, he said, to figure out what factors are important in colon cancer development.

Title: **France reports leveling childhood obesity rates**

Publish Date: **2008-05-15**

http://feeds.reuters.com/~r/reuters/healthNews/~3/290957211/idUSL1591757720080515

France is the first EU country to report a leveling off of childhood obesity rates, suggesting that healthier diet programs and a ban on vending machines in schools is paying off, researchers said on Thursday.

The findings from two separate studies of school-age children signal a shift in France after decades of increase, researchers told the 2008 European Congress on Obesity.

"The rates of children who are overweight are undergoing an overall stabilization in France among all socio-economic backgrounds," said Sandrine Lioret, an epidemiologist at the French Food Safety Agency, who led one of the studies.

Obesity is a major problem worldwide that increases the risk later in life of type 2 diabetes, cancer and heart disease. The World Health Organization classifies around 400 million people as obese, including 20 million children under age 5.

Many Western governments -- where the obesity problem is greatest -- have adopted programs in recent years to promote healthier diets and lifestyles to keep children from growing up to be overweight and obese.

The French findings are important because they show that government policies are a potential weapon in the fight against childhood obesity, said Tim Lobstein, a director of the International Obesity Task Force in London.

But he cautioned that only time would tell whether the French results are a one-off blip or part of a long-term trend.

"The tidal wave (of obesity) is continuing to surge in most European countries," he said. "We are seeing that wave roll on through to adulthood."

France is about in the middle when it comes to European childhood obesity rates, with the lowest seen in Scandinavia and the highest in poorer nations in Southern Europe, the researchers said.

In one of the studies, Lioret's team at the French Food Safety Agency showed no statistically significant change in the prevalence of obesity rates among randomly-selected school children age 3 to 17 in surveys taken eight years apart.

The other research from the French National Institute for Health Surveillance found that the number of obese children aged 7 to 9 had remained steady at around 18 percent in 2000 and 2007.

Government policies, a growing awareness of the dangers of obesity and the fact that children are eating less all seem to be playing a role, the researchers said.

One worry, though, is that even as the overall rate has flattened, poor children were up to three times more likely to be obese compared to wealthier children, they added.

Title: **Weight-loss surgery won't "cure" sleep apnea**

Publish Date: **2008-08-29**

http://feeds.reuters.com/~r/reuters/healthNews/~3/378275902/idUSHAR96402420080829

In people who are obese, weight-loss surgery will likely lead to an improvement in obstructive sleep apnea (OSA) but it won't eliminate the nighttime breathing disorder. Many patients will have residual OSA one year after weight-loss surgery (also known as bariatric surgery), results of a study indicate.

"There are numerous benefits to weight loss by any means, (including) a reduction in the severity of OSA," study leader Dr. Christopher Lettieri of Walter Reed Army Medical Center in Washington, DC, told Reuters Health. "However, patients and their physicians should understand that OSA can occur in the absence of obesity, and losing weight, even if substantial, may not resolve OSA."

OSA is a common problem, particularly among the obese, in which tissues in the back of the throat temporarily collapse during sleep causing numerous, brief episodes of interrupted breathing. It can be effectively treated with a special "CPAP" breathing device that alleviates the blockage by pushing air into throat.

In a study designed to clarify the impact of bariatric surgery on OSA, 24 morbidly obese patients underwent overnight sleep studies before and 1 year after bariatric surgery.

All of them had OSA at the start of the study and surgical weight loss resulted in substantial improvements in the severity of OSA, Lettieri and colleagues report in the Journal of Clinical Sleep Medicine.

However, all but one patient had persistent OSA despite their weight loss. "In fact, the majority still had moderate to severe disease, which would require continued treatment," Lettieri said. Two people had a worsening of their OSA despite significant weight loss.

"OSA," Lettieri said, "is associated with numerous adverse effects on health and quality of life, especially in those with moderate to severe disease. If present, it should be treated."

Patients having weight-loss surgery, he added, should not assume their OSA has resolved and should have a repeat sleep study prior to discontinuing their OSA treatment.

Title: **Pelvic disorders common among women: study**

Publish Date: **2008-09-16**

http://feeds.reuters.com/~r/reuters/healthNews/~3/394705319/idUSN1642412220080917

Nearly one quarter of all U.S. women have some sort of pelvic floor disorder such as urinary or fecal incontinence, and more cases are likely as the population ages, researchers said on Tuesday.

Urinary incontinence or loss of bladder control is by far the most common problem, and childbirth, which can weaken pelvic floor muscles, is the most frequent cause.

"What stands out is such a large number of women had symptoms of a moderate-to-severe pelvic floor disorder," said Dr. Ingrid Nygaard of the University of Utah, whose study appears in Journal of the American Medical Association.

"We know many women leak urine once in a blue moon if they are doing strenuous exercise. That such a large number leaks on a regular basis was surprising to see," she said by telephone.

Nygaard's team studied 1,961 nonpregnant women aged 20 and older who were part of a national health survey in 2005-2006. They were examined for symptoms of urinary and fecal incontinence and pelvic organ prolapse, in which an organ such as the bladder drops and pushes against the walls of the vagina.

Overall, 23.7 percent of women reported symptoms of at least one pelvic floor disorder. Of these, 15.7 percent had urinary incontinence, 9 percent had fecal incontinence and 2.9

percent had pelvic organ prolapse, in which women reported seeing or feeling a vaginal bulge.

Age plays a major role, with just 10 percent of women 20 to 39 reporting at least one disorder, compared with about a quarter of women 40 to 59, 37 percent of women 60 to 79, and nearly half of women 80 and older.

Women were considered to have urinary incontinence if they had at least weekly leakage, or monthly leakage of volumes of more than just a few drops.

If women with occasional leakage were included, the numbers might be much higher, the researchers said.

"The study results underscore the need to identify the causes of pelvic floor disorders and the means to prevent and treat them," Dr. Duane Alexander, director of the National Institute of Child Health and Human Development, said in a statement.

Treatments for urinary stress incontinence -- the kind linked with laughing, coughing, sneezing or exercise -- range from exercises to surgical options.

Urge incontinence or overactive bladder can be treated by lifestyle changes and a number of drugs, including tolterodine or Detrol made by Pfizer Inc; Vesicare or solifenacin made by GlaxoSmithKline and Astellas Pharma Inc; and Enablex or darifenacin by Novartis AG.

Title: **Exercise and limited TV time may keep kids trim**

Publish Date: **2008-08-22**

http://feeds.reuters.com/~r/reuters/healthNews/~3/371961724/idUSPAT25502520080822

Children who regularly exercise and limit their time in front of the TV and computer are much less likely to be overweight than their peers, a new study suggests.

The American Academy of Pediatrics recommends that children devote no more than two hours per day to watching TV and playing video games.

Experts also encourage children to exercise regularly; some groups, including the AAP, recommend that boys move enough to take 13,000 steps each day, while girls should strive for 11,000. Another common recommendation is for children and teenagers to get at least one hour of moderate exercise on most, if not all, days of the week.

For the new study, researchers at Iowa State University in Ames looked at whether there were weight differences between children who met or did not meet recommendations for "screen time" and exercise.

They found that among 709 7- to 12-year-olds, those who did not meet either recommendation were three to four times more likely to be overweight than their peers who met both guidelines.

The findings are published in the Journal of Pediatrics.

Since the study compared groups of children at one point in time, it does not prove that following exercise and screen-time recommendations keeps children at a healthy weight, according to lead researcher Dr. Kelly R. Laurson.

However, the findings do show that screen time and exercise are each independently associated with the odds of a child being overweight, explained Laurson, who is now with Illinois State University in Normal.

This gives support to experts' guidelines, the researcher told Reuters Health, and "parents should encourage their children to comply with both."

Laurson and colleagues based their findings on 709 children from 10 schools in Iowa and Minnesota. They surveyed the children about their TV watching and video game habits, and they had each child wear a pedometer, or step counter, for at least four days in order to gauge his or her typical physical activity level.

Boys who took at least 13,000 steps per day were considered to be in line with exercise recommendations, as were girls who logged 11,000 or more steps.

The researchers found that among boys who met the recommendations on exercise and screen time, 10 percent were overweight. Among girls meeting both recommendations, 20 percent were overweight. That compared with 35 percent of boys and 40 percent of girls who met neither recommendation.

"Both physical activity and screen time are important factors in childhood overweight," Laurson said.

Unfortunately, the investigators found, only a minority of children in their study were meeting experts' recommendations. About 44 percent were getting enough exercise, while just 31 percent were limiting their screen time to less than two hours per day.

Title: **Nature tops nurture in childhood obesity: study**

Publish Date: **2008-02-07**

http://feeds.reuters.com/~r/reuters/healthNews/~3/231131326/idUSL0793070220080207

Diet and lifestyle play a far smaller role than genetic factors in determining whether a child becomes overweight, according to a British study of twins published on Thursday.

Researchers looking at more than 5,000 pairs of twins wrote in the American Journal of Clinical Nutrition that genes account for about three-quarters of the differences in a child's waistline and weight.

"Contrary to the widespread assumption that family environment is the key factor in determining weight gain, we found this was not the case," said Jane Wardle, director of Cancer Research UK's Health Behavior Centre, who led the study.

Previous studies have pointed to environmental factors as the main cause of obesity, a major problem worldwide that increases the risk later in life of type-2 diabetes, cancer and heart problems.

The World Health Organization classifies around 400 million people worldwide as obese, including 200 million children under the age of five.

The British team looked at pairs of identical twins who share all their genes and compared their measurements with those of non-identical twins who share only half their genes.

A statistical analysis found that the differences in the children's body mass index and waist circumference were 77 percent attributable to genes and 23 percent due to the environment in which the children were growing up.

BMI is calculated by dividing weight by the square of height.

"These results do not mean that a child with a high complement of 'susceptibility genes' will inevitably become overweight, but that their genetic endowment gives them a stronger predisposition," the researchers said.

The results suggest that parents whose children are at the greatest genetic risk may need support to make sure they provide a healthy environment, the researchers said.

"This study shows that it is wrong to place all the blame for a child's excessive weight gain on the parents," the researchers said.

Title: **Hormone discovery may help combat diabetes: study**

Publish Date: **2008-09-18**

http://feeds.reuters.com/~r/reuters/healthNews/~3/396511401/idUSN1842268420080918

Scientists have identified a member of a new class of hormones produced by body fat that they think could lead to fresh approaches to combat diabetes and other conditions related to obesity.

The hormone prevents the liver from accumulating fat and enhances the body's ability to control glucose, scientists led by Gokhan Hotamisligil of the Harvard School of Public Health wrote on Thursday in the journal Cell.

Their work involving the hormone, called palmitoleate, was in mice, but the hormone is also found in people.

While other known hormones are either protein-based or steroid-based, this one is the first example of a class made out of fatty acids, Hotamisligil said. The researchers are calling this new group of hormones lipokines.

Hormones act like the body's chemical messengers, traveling through the blood to influence cells and organs in processes such as growth and development, metabolism, sex and mood.

If the hormone's role in people is the same as in mice, it may become a valuable weapon against type 2 diabetes or fatty liver disease, Hotamisligil said.

Scientists previously had known about palmitoleate but had not identified it as a hormone, he said.

"All evidence is pointing that it's coming from fat cells," Hotamisligil said in a telephone interview.

One of its roles is to communicate with the liver and prevent it from accumulating fat, which can occur as people become obese, he said. It also encourages muscle to take up glucose from blood and dispose of it, he added.

It works almost as well as the hormone insulin at pushing sugar out of the blood, Hotamisligil said. Insulin regulates the absorption of sugar into the cells.

People with diabetes have blood sugar levels that are too high. Those with type 2 diabetes, the form closely related to obesity, are resistant to the effects of insulin or produce too little of it.

The researchers said that as body fat increases, less palmitoleate is produced. So in obese people, the beneficial functions of this hormone in controlling blood sugar levels and preventing fat accumulating in the liver would be diminished.

"When you need it the most, you produce the least," Hotamisligil said.

Doctors potentially could give palmitoleate to people or come up with ways to stimulate the body to produce more to prevent or improve illnesses like diabetes, Hotamisligil said.

And a simple test looking at blood levels of palmitoleate potentially could be used to signal risk for conditions like diabetes.

The scientists identified palmitoleate as a hormone with the help of scientists from Lipomics Technologies of West Sacramento, California. The company was acquired on Wednesday by Tethys Bioscience Inc of Emeryville, California.

Title: **Unilever says new milkshake helps control appetite**

Publish Date: **2008-05-14**

http://feeds.reuters.com/~r/reuters/healthNews/~3/290286237/idUSL1484767120080514

A new weight loss drink that tastes like a milkshake significantly reduces appetite and could soon join Unilever's $400 million Slim-Fast weight-loss brand, the company's researchers said on Wednesday.

A study showed that the drink, which works by trapping gas in foods to make people feel full, worked even better than the company's Slim-Fast weight-loss drink, they said.

The researchers, who presented their findings at the 2008 European Congress on Obesity, said the company has patented the technology.

"The technology is now available for the brand to use in future formats," said David Mela, a Unilever nutritionist who worked on the study. "The food maintains the bulk, much of which is air that helps you maintain that full feeling."

Obesity is a big problem and big business. About 400 million people are classified as obese, putting them at higher risk of diseases like type 2 diabetes and heart diseases, according to the World Health Organization.

This has in part spurred companies like Unilever, Kraft Foods, General Mills, Sara Lee and others to turn to healthier products with a whole range of so-called health and wellness foods.

"If you look at western populations, a (large number) of adults are overweight," said Gert Meijer, an executive at Unilever's research and development division. "In terms of the amount of people who might be interested in this product, it could be huge."

In the Unilever study, the researchers tested their milkshake on 24 volunteers who were given either the new drink or a serving of regular Slim-Fast at breakfast.

People who had the milkshake reported that they were significantly fuller when asked at different intervals over a four-hour time period. The researchers found that a half-sized serving of the milkshake also suppressed hunger.

"We are clearly talking about hours," said Sergei Melnikov, a physical chemist who helped develop the technology. "It is an effect that lasts for an hour or two or longer."

The milkshake is designed to trap gas in the food after consumption, preventing it from dissolving in the mouth as happens with foods like whipped cream, and cutting appetite.

To do this the team engineered the fats, proteins and fibers in the food until reaching the right mix to trap the gas -- a technology that might appear in other Unilever foods, the researchers said.

"I would say this is not limited to liquids," Melnikov said. "It could be used in other food forms."

Title: **Obesity may diminish a man's fertility**

Publish Date: **2008-09-19**

http://feeds.reuters.com/~r/reuters/healthNews/~3/397526706/idUSKEN97315720080919

Being obese may dim a man's chances of becoming a father, even if he is otherwise healthy, a new study suggests.

Researchers found that among 87 healthy men ages 19 to 48, those who were obese were less likely to have ever fathered a child. More importantly, they showed hormonal differences that point to a reduced reproductive capacity, the researchers report in the journal Fertility and Sterility.

Compared with their thinner counterparts, obese men had lower levels of testosterone in their blood, as well as lower levels of luteinizing hormone (LH) and follicle-stimulating hormone (FSH) -- both essential to reproduction.

According to the researchers, these relatively low levels of LH and FSH are suggestive of a "partial" hypogonadotropic hypogonadism. This is a condition in which the testes do not function properly due to signaling problems in the hypothalamus or pituitary gland, two brain structures involved in hormone secretion.

The findings suggest that obesity alone is an "infertility factor" in otherwise healthy men, write Dr. Eric M. Pauli and his colleagues at the Pennsylvania State University College of Medicine in Hershey.

Of the 87 men in the study, 68 percent had had a child. Pauli's team found that the average body mass index, or BMI, was lower among these men compared with those who'd never fathered a child; in the former group, the average BMI was 28, which falls into the range for "overweight," while the average BMI for childless men was nearly 32, which falls into the "obese" range.

When the researchers assessed the men for several reproductive hormones, they found that the more obese a man was, the lower was his LH and FSH levels. On the other hand, increasing obesity correlated with increasing estrogen levels.

Excess body fat, Pauli's team explains, may increase the conversion of testosterone to estrogen in a man's blood. Such hormone alterations could, in turn, signal the brain to suppress FSH and LH production.

Past studies have linked obesity with a dampened libido and increased risk of erectile dysfunction, the researchers note. Those effects, they say, along with the hormonal alterations seen in this study, could act together to decrease an obese man's fertility.

Title: **Lower thyroid activity tied to weight gain**

Publish Date: **2008-04-04**

http://feeds.reuters.com/~r/reuters/healthNews/~3/264169549/idUSCOL46971820080404

Middle-aged adults whose thyroid gland is mildly underactive, but still functioning in the normal range, may be more prone to weight gain, a new study suggests.

The thyroid is a gland in the neck that produces hormones that regulate the body's metabolism. In a disorder called hypothyroidism, the gland is underactive, causing symptoms such as fatigue, sensitivity to cold, dry skin and weight gain.

But it has been unclear whether thyroid function within the standard range has an effect on body weight.

In the current study, reported in the Archives of Internal Medicine, researchers looked at the relationship between body weight and levels of thyroid-stimulating hormone (TSH) in more than 2,400 middle-aged adults.

TSH is released by the brain to stimulate hormone production in the thyroid gland. Higher TSH levels in the blood indicate relatively lower activity in the thyroid.

In this study, men and women with relatively high, but still normal, TSH levels tended to weigh more at the outset than those with lower TSH concentrations.

Moreover, those whose TSH levels tipped upward over the next several years were more prone to weight gain.

"Our findings raise the possibility that modest increases in serum TSH concentrations within the reference range may be associated with weight gain," write the researchers, led by Dr. Caroline S. Fox of the National Heart, Lung, and Blood Institute in Bethesda, Maryland.

It's too soon, however, to start tinkering with thyroid hormones in order to treat obesity, editorialists comment.

Metabolism is governed by a complex interaction between the nervous system and hormone-producing glands. And while this system, including thyroid hormones, may influence weight and obesity risk, obesity also seems to affect the system, according to Drs. Roy E. Weiss and Rebecca L. Brown of the University of Chicago Medical Center.

Several studies, they note, have shown that excess fat tissue might directly affect TSH levels.

The study included 2,407 men and women who were an average of 48 years old when the study began. Among the women, the average weight for those with the lowest TSH levels was 142 pounds, versus 155 among those with the highest TSH levels; the corresponding figures for men were 182 pounds and 189 pounds.

Over the next 3.5 years, the group as a whole put on a few pounds. However, men and women whose TSH levels crept up tended to gain more.

Women with the highest TSH levels gained an average of 9.3 pounds more than women with the lowest TSH levels. The average weight gain in men with the highest TSH levels compared with those with the lowest levels was 4.2 pounds greater.

More research, according to Fox's team, is needed to confirm the findings, and to understand why TSH levels are connected to weight.

Title: **Strict Mediterranean diet offers big health boost**

Publish Date: **2008-09-11**

http://feeds.reuters.com/~r/reuters/healthNews/~3/390125707/idUSLB73099820080911

Sticking strictly to a Mediterranean diet rich in fruits and vegetables offers substantial protection against cancer, heart disease and other major chronic illnesses, Italian researchers said on Friday.

People who did this had a 9 percent drop in death from heart disease, a 13 percent reduction in incidence of Parkinson's and Alzheimer's disease and a 6 percent reduction in cancer compared to those who were not as diligent, their study found.

"These results seem to be clinically relevant for public health, particularly for encouraging a Mediterranean-like dietary pattern for primary prevention of major chronic diseases," wrote Francesco Sofi, a clinical nutrition researcher, and colleagues at the University of Florence.

The traditional Mediterranean diet is full of vegetables, fish and healthy fats such as olive oil, and low in red meat, dairy products and alcohol.

Sofi and his team reviewed 12 international studies which included more than 1.5 million people whose eating habits and health were tracked for follow-up periods of three to 18 years.

The researchers also developed an "adherence" score to rate how well people followed the Mediterranean diet, a tool they said doctors could use to help improve people's health and encourage them to eat better.

"The adherence score...could be an effective preventative tool for reducing the risk of mortality and morbidity in the general population," they wrote.

Title: **Multivitamins are top diet supplement for teens**

Publish Date: **2008-04-18**

http://feeds.reuters.com/~r/reuters/healthNews/~3/273033791/idUSCOL86248220080418

A new study indicates that multivitamins and vitamin C top the list of dietary supplements used by US adolescents, which is "reassuring" given the relative lack of health risks associated with them, researchers say.

But adolescents in the study who used prescription medications were also more likely to use dietary supplements, and doctors and pharmacists should be sure to ask their young patients about supplements to avoid the possibility of harmful interactions, Dr. Paula Gardiner of Boston University Medical School and her colleagues conclude.

Gardiner and her team reviewed data from the 1999-2002 National Health and Nutrition Examination Surveys for 11- to 19-year-olds to investigate how common supplement use is among adolescents and factors associated with using vitamins, herbal medicines, minerals and other products.

Twenty-seven percent of the adolescents surveyed said they had used a dietary supplement in the past month, the researchers found. Sixteen percent used multivitamins, while 6 percent said they took vitamin C. Just 4 percent used non-vitamin mineral supplements, including 2 percent who said they used supplements to help them lose weight or enhance sports performance.

Non-Hispanic whites were most likely to be using dietary supplements, while prescription medication users were 37 percent more likely than those not taking prescribed drugs to use dietary supplements. Study participants who said they were in fair or poor health were 41 percent less likely to take supplements than their peers who considered themselves to be in better health. And adolescents who reported having chronic headaches were 25 percent more likely to use dietary supplements.

Obese individuals were 51 percent more likely to be using non-vitamin or mineral herbal supplements, the researchers found, as were older teens.

"To better understand use among culturally diverse groups and those with different clinical conditions, future studies should include a broader range of dietary supplements (such as those used in folk remedies, foods and medicinal teas) and ask about common health conditions," the researchers conclude.

"Additional studies are needed to determine the impact of dietary supplement use on health care use, health status, and quality of life," they add.

Title: **Reduced "exercise capacity" an ominous sign**

Publish Date: **2008-02-07**

http://feeds.reuters.com/~r/reuters/healthNews/~3/231179397/idUSCOL77254720080207

People who have trouble exercising on a treadmill are at increased risk of suffering a heart attack or other heart-related event and of dying, according to results of a study.

"Exercise capacity" is one of many important prognostic factors measured during an exercise treadmill test, a simple procedure often performed in doctors' offices. But "little is known about the association between exercise capacity among patients referred for exercise treadmill testing and nonfatal cardiac events," Dr. Pamela N. Peterson, of the University of Colorado at Denver and Health Sciences Center, and colleagues note in a report.

Among 9191 adults who had a treadmill test and who were followed for a median of 2.7 years, 119 were hospitalized for heart attack and 259 for chest pain. Moreover, 749 required revascularization procedures to restore blood flow to the heart, and 132 patients died.

According to Peterson's team, people with low exercise capacity, relative to those with normal exercise capacity, on the treadmill test, had more than a twofold increased risk of having a heart attack, experiencing chest pain, or needing a revascularization procedure.

Low exercise capacity was also significantly associated with an increased risk of dying from any cause.

Peterson's team also found that individuals with lower exercise capacity were more likely to be female and to have other comorbid conditions such as diabetes and high blood pressure compared to individuals with greater exercise capacity.

Those with lower exercise capacity were also more likely to have chest pain on the treadmill and abnormal heart rate recovery -- a measure of how quickly the heart rate returns to normal after a period of exercise.

The findings have important implications, the researchers conclude, suggesting that "aggressive risk factor modification and close follow-up should be considered for patients with impaired exercise capacity."

Title: **Healthy diet means better school performance**

Publish Date: **2008-04-14**

http://feeds.reuters.com/~r/reuters/healthNews/~3/270229893/idUSTON47353620080414

Kids who eat better perform better in school, a new study of Nova Scotia fifth-graders confirms.

Students who ate an adequate amount of fruit, vegetables, protein, fiber and other components of a healthy diet were significantly less likely to fail a literacy test, Dr. Paul J. Veugelers of the University of Alberta in Edmonton and colleagues found.

While a healthy diet is generally assumed to be important for good school performance, there has actually been little research on this topic, Veugelers and his colleagues note. To investigate, they looked at 4,589 fifth-graders participating in the Children's Lifestyle and School-performance Study, 875 (19.1 percent) of whom had failed an elementary literacy assessment.

The better a student's eating habits based on several measures of diet quality, including adequacy and variety, the less likely he or she was to have failed the test, the researchers found, even after they adjusted the data for the effects of parental income and education, school, and sex. Eating plenty of fruit and vegetables, and getting fewer calories from fat, was also associated with a lower risk of failing the test.

To date, Veugelers and his team say, most research on diet and school performance has focused on the importance of eating breakfast, as well as the ill effects of hunger and malnutrition.

"This study extends current knowledge in this area by demonstrating the independent importance of overall diet quality to academic performance," the researchers conclude.

"The consistency of this association across various indicators of diet quality gives emphasis to the importance of children's nutrition not only at breakfast but throughout the day."

Title: **Nonfat milk linked to prostate cancer**

Publish Date: **2008-01-02**

http://feeds.reuters.com/~r/reuters/healthNews/~3/210044591/idUSHAR26781420080102

The amount of calcium and vitamin D in the diet appears to have little or no impact on the risk of prostate cancer, but the consumption of low-fat or nonfat milk may increase the risk of the malignancy, according to the results of two studies published in the American Journal of Epidemiology.

Dietary calcium and dairy products have been thought to increase the risk of prostate cancer by affecting vitamin D metabolism. Data from several prospective studies have supported an association, but many other studies have failed to establish a link.

To explore this topic further, Dr. Song-Yi Park, from the University of Hawaii in Honolulu, and colleagues, analyzed data from subjects enrolled in the Multiethnic Cohort Study. This study, conducted between 1993 and 2002, included adults between 45 and 75 years old, were primarily from five different ethnic or racial groups, and lived in California or Hawaii.

A total of 82,483 men from the study completed a quantitative food frequency questionnaire and various factors, such as weight, smoking status, and education levels were also noted, Park's group said.

During an average follow-up period of 8 years, 4,404 men developed prostate cancer. There was no evidence that calcium or vitamin D from any source increased the risk of prostate cancer. This held true across all racial and ethnic groups.

In an overall analysis of food groups, the consumption of dairy products and milk were not associated with prostate cancer risk, the authors found. Further analysis, however, suggested that low-fat or nonfat milk did increase the risk of localized tumors or non-aggressive tumors, while whole milk decreased this risk.

In a similar analysis, Dr. Yikyung Park, from the National Cancer Institute at National Institutes (NIH) of Health in Bethesda, Maryland, and colleagues investigated the relationship of calcium and vitamin D and prostate cancer in 293,888 men enrolled in the NIH-American Association of Retired Persons Diet and Health Study, conducted between 1995 and 2001. The average follow-up period was 6 years.

No link between total or supplemental dietary calcium and the total number of non-advanced prostate cancer cases was noted. Total calcium intake was tied to advanced and fatal disease, but both associations fell short of statistical significance.

Similar to the first study's findings, skim milk was linked with advanced prostate cancer. Calcium from non-dairy food, by contrast, was tied to a reduced risk of non-advanced prostate cancer.

"Our findings do not provide strong support for the hypothesis that calcium and dairy foods increase the risk of prostate cancer. The results from other large...studies, with adequate numbers of advanced and fatal prostate cancers, may shed further light on this question," Park's team concludes.

Title: **What a nightmare: Americans get too little sleep**

Publish Date: **2008-02-28**

http://feeds.reuters.com/~r/reuters/healthNews/~3/243048146/idUSN2859197720080229

With late-night TV watching, Internet surfing and other distractions, Americans are getting less and less sleep, the U.S. Centers for Disease Control and Prevention said on Thursday.

And all this sleeplessness can be a nightmare for your mental and physical health, CDC experts cautioned, calling sleep loss an under-recognized public health problem.

Sleep experts say chronic sleep loss is associated with obesity, diabetes, high blood pressure, stroke, cardiovascular disease, depression, cigarette smoking and excessive drinking.

The CDC surveyed 19,589 adults in four states. Ten percent reported they did not get enough sleep or rest every single day of the prior month, and 38 percent said they did not get enough in seven or more days in the prior month.

The CDC survey was conducted in New York, Hawaii, Delaware and Rhode Island, asking people how many days in the prior month they got insufficient rest or sleep, without asking specifically how many hours they slept.

But the CDC released nationwide data collected separately showing that across all age groups, the percentage of adults reporting sleeping six hours or fewer a night increased from 1985 to 2006.

The National Sleep Foundation recommends adults get seven to nine hours of sleep a night. Children ages 5 to 12 should get nine to 11 hours and those 11 to 17 need 8-1/2 to 9-1/2 hours.

SLEEP IS VITAL

"At night, we're doing everything except for sleeping -- we're on the Internet, we may be watching TV. With these new lifestyles we have kind of taken sleep for granted as something that we can do when we have time or we can catch up on it on the weekends," CDC behavioral scientist Lela McKnight-Eily, who led the study, said in a telephone interview.

"We don't realize that sleep is a vital part of overall health and that chronic sleep loss is related to both physical and mental health issues," she added. "It's getting worse."

Darrel Drobnich, National Sleep Foundation chief executive officer, added that several thousand people die on U.S. roads yearly in accidents involving drowsy drivers.

"Americans are definitely sleep deprived. They don't get the amount that even they say that they want," Drobnich said.

The CDC said 50 to 70 million Americans suffer from chronic sleep loss and sleep disorders in a country of 300 million.

The CDC four-state survey found that younger adults are more likely than older adults to report getting too little sleep. It also found overall that 30 percent of respondents said they got enough sleep every day of the past month, and 33 percent got too little on one to six days in the prior month.

Lela McKnight-Eily urged people who often get too little sleep to see a doctor to see whether lifestyle issues are to blame or whether they might have a sleeping disorder. People can also try to establish a regular sleep schedule and avoid caffeine or other stimulants before bedtime, she added.

Title: **Preterm babies at risk of hospitalization as adults**

Publish Date: **2008-01-03**

http://feeds.reuters.com/~r/reuters/healthNews/~3/210754870/idUSLAU37846020080103

People who were born early or just had an unusually low birth weight are more likely to be hospitalized in adolescence and young adulthood, Swedish researchers report.

Being small for gestational age (SGA) -- significantly smaller than most babies born after the same number of weeks of pregnancy -- was a greater risk factor for future hospitalization than preterm birth, Dr. Katarina Ekholm Selling of Linkoping University and her colleagues found.

While the early-life health consequences of SGA and preterm birth have been investigated extensively, there has been less research on how these individuals fare later in life, Selling noted in an interview with Reuters Health.

To investigate, she and her team looked at the hospitalization records between 1987 and 1996 for every person born in Sweden between 1973 and 1975, a total of 304,275 men and women.

Overall, SGA men and women were 16 percent more likely to be hospitalized than individuals who were born at a normal size, while having been born preterm increased hospitalization risk by 6 percent. People who had been both SGA and preterm were 42 percent more likely to be hospitalized.

Hospitalizations for mental disorders, drug use, injuries and poisoning, as well as "symptoms, signs and ill-defined conditions," poorly defined intestinal infections and genitourinary diseases, were more frequent for SGA individuals.

Those born preterm were more likely than full-term infants to be hospitalized for endocrine, nutritional and metabolic diseases, mental disorders and nervous system diseases, birth defects and "symptoms, signs and ill-defined conditions."

While the study was not designed to show the mechanism behind the increased risk of hospitalization for SGA and preterm individuals, Selling said, the greater likelihood of accidents, drug use and mental health problems seen with SGA suggest that "personality may be the key."

One possibility could be that SGA individuals are more prone to risk-taking behavior, she added, but studies investigating the relationship have had conflicting results.

SGA and preterm individuals also tend to be from more disadvantaged socioeconomic backgrounds, Selling said, and while her team attempted to use statistical techniques to control for the effect of socioeconomic status, it may still be involved in the relationship.

Nevertheless, she pointed out, "there are many other factors that are far more important in determining our health in later life than being born small for gestational age."

Title: **Laparoscopic gastric bypass provides better results**

Publish Date: **2008-07-15**

http://feeds.reuters.com/~r/reuters/healthNews/~3/336316634/idUSCOL56557220080715

Performing gastric bypass surgery to reduce the weight of morbidly obese patients using a laparoscopic method, rather than the conventional more invasive "open" abdominal method, reduces postoperative complications, the need for a second operation, and shortens hospital stays, new research shows. Nevertheless, laparoscopic gastric bypass is more expensive.

Obesity surgery, also called bariatric surgery, is growing in popularity and more and more of these operations are being done using a laparoscope, note co-authors Dr. Wendy E. Weller, from the University at Albany in New York, and Dr. Carl Rosati, from Albany Medical Center.

This is done by placing one or more small incisions in the abdomen, through which a hollow tube is inserted. This allows very small instruments to be inserted to perform the gastric bypass. The entire procedure is visualized on a screen. In contrast, the more invasive "open" procedure involves making an incision to open the abdomen so the procedure can be performed.

The current study, reported in the Annals of Surgery, involved an analysis of data from 19,156 subjects who underwent gastric bypass surgery in 2005 and were logged in the Nationwide Inpatient Sample, the largest all-payer inpatient database in the U.S.

Slightly less than 75 percent of the patients underwent laparoscopic gastric bypass, the report indicates.

Laparoscopic gastric bypass was linked to a reduced risk of several complications. With open surgery, the risk of pulmonary complications was increased by 92 percent, for cardiovascular complications it was 54 percent, for sepsis, a serious system-wide infection, the risk was more than doubled and the risk of anastomotic leak, leakage from the operative site, 32 percent higher.

On average, performing laparoscopic rather than open gastric bypass reduced the hospital stay by about 1 day.

The average total charges were similar for the two procedures, but median total charges were significantly higher with laparoscopic gastric bypass: $30,033 vs. $28,107 respectively.

After accounting for various patient and hospital factors, laparoscopic surgical patients were less likely than their open-surgery counterparts to require reoperation, the investigators found.

While these findings suggest some advantages with the laparoscopic operation, "most reassuring for the bariatric surgery community is that the hospital outcomes were excellent overall in both the laparoscopic and open procedures," Dr. Michael G. Sarr, from the Mayo Clinic in Rochester, Minnesota, comments in a related editorial.

Title: **Exercise plus relaxation may lessen migraine pain**

Publish Date: **2008-07-31**

http://feeds.reuters.com/~r/reuters/healthNews/~3/351748556/idUSCOL15942120080731

A combination of aerobic exercise and muscle relaxation may help ease migraine pain, a small study suggests.

Austrian researchers found that when 15 migraine sufferers added an exercise-and-relaxation routine to their usual care, the patients reported an improvement in migraine pain intensity over six weeks.

The findings, reported in the Clinical Journal of Sports Medicine, add to evidence that exercise may offer some relief from migraine pain. A recent research review, for instance, concluded that while it seems unlikely that exercise prevents migraine attacks, it may make them less intense when they do arise.

For the current study, Dr. Martin Kopp and colleagues at Innsbrook Medical University randomly assigned 30 female migraine patients to one of two groups. Half stayed with their usual treatment only, while the other half added a twice-weekly exercise routine -- 45 minutes of aerobic exercise and 15 minutes of progressive muscle relaxation.

After six weeks, women in the exercise group reported a greater improvement in migraine pain intensity than their counterparts in the comparison group.

The exercisers also reported fewer depression symptoms at the end of the study, though overall, their psychological well-being was no different from that of women in the comparison group.

The results leave several important questions, according to Kopp's team. One is whether the exercise, relaxation or both were responsible for the pain improvements.

Another question is why women in exercise/relaxation group reported lesser pain intensity. It's possible, Kopp and his colleagues note, that physical activity, relaxation and other non-drug migraine treatments all foster feelings of self-efficacy, which can help people cope with pain.

More studies with larger groups of patients are needed to understand how non-medication migraine therapies work, and which ones are most effective, they conclude.

Title: **Survey gauges side effects of prostate treatments**

Publish Date: **2008-03-19**

http://feeds.reuters.com/~r/reuters/healthNews/~3/254525122/idUSN1932154520080319

Age, race and obesity affect how satisfied men are with their treatment for prostate cancer, U.S. researchers said on Wednesday.

And the effects of short-term hormone therapy can linger for years, the survey of 1,201 men treated at nine university hospitals and 625 of their partners found.

The results, published in the New England Journal of Medicine, are designed to give doctors and patients a better idea of what to expect from three types of prostate cancer treatment.

"I don't think our findings are going to make any one of the specific approaches a winner," said Dr. Martin Sanda of Beth Israel Deaconess Medical Center in Boston.

"But they do make it possible for doctors and patients to better gauge what to expect for treatment A or treatment B," Sanda said.

One year after treatment, sexual functioning was a moderate or big problem among 50 percent of men whose prostate had been removed, the researchers found.

It was a problem for 31 percent treated with external radiation and 30 percent who had radioactive seeds placed in the prostate. Surgery to spare the nerves, they found, helped prevent sexual problems.

"Overall, 10 to 19 percent of patients or their partners reported being distressed by symptoms attributable to hormonal therapy," the researchers wrote, adding that the finding raises questions about whether hormone therapy should be restricted to high-risk cases.

Therapy designed to block the male hormone testosterone for six months had effects on sexuality and vitality that persisted for up to two years, the doctors found. Such treatments are usually combined with nonsurgical

therapy.

RACE A MAJOR FACTOR

"The one factor that really stood out across all the different treatments was that African-American men were less satisfied with their outcome," said Sanda.

"It could be differences in expectations, if doctors are not communicating as effectively. Another possibility is that the aggressiveness of prostate cancer tends to be worse for them," Sanda added.

"I think a lot of unhappiness that people experience with prostate cancer may be because the specific information about what will happen has not been all that available in a way that is really relevant to their concerns, which differ from patient to patient," he said.

Urinary incontinence was a moderate or big problem in 8 percent of surgery patients, 5 percent of seed patients and 4 percent of radiotherapy patients.

Title: **Worsening incontinence not linked to menopause**

Publish Date: **2008-03-27**

http://feeds.reuters.com/~r/reuters/healthNews/~3/259059047/idUSCOL75650520080327

The worsening of symptoms of incontinence among middle-aged women is attributable to weight gain, not menopause, according to findings published in the journal Obstetrics and Gynecology.

Previous studies have found a higher rate of urinary incontinence in women ages 45 to 55 years, coinciding with the menopause transition, note Dr. L. Elaine Waetjen, of the University of California, Davis, and colleagues. This increase in incontinence in midlife has been explained, in part, by urinary tract changes associated with the loss of estrogen during menopause.

The researchers examined the validity of this explanation by analyzing data from 2,415 women who reported episodes of incontinence monthly or more often at study enrollment and during the first six annual follow-up visits (1995 to 2002) of the Study of Women's Health Across the Nation.

Worsening incontinence was defined as an increase in the frequency and improving incontinence was defined as a decrease in the frequency between annual visits. Questions assessing menstrual bleeding patterns were used to classify the menopausal status annually of the women not on hormone therapy.

Overall, 14.7 percent of incontinent women reported worsening incontinence over 6 years. Improvement was reported by 32.4 percent of women, and no change in the frequency of symptoms was reported by 52.9 percent.

As noted, the transition through menopause did not significantly affect the severity of incontinence symptoms. By contrast, weight gain was associated with worsening incontinence.

"Many women and clinicians have believed urinary incontinence to be a symptom attributable to the menopausal transition, but our results suggest that the transition...has either no effect or possibly a weak positive effect on changes in the frequency of incontinence symptoms in midlife women," Waetjen and colleagues conclude.

Title: **Lawmakers probe FDA approval of Ranbaxy drugs**

Publish Date: **2008-07-22**

http://feeds.reuters.com/~r/reuters/healthNews/~3/343163559/idUSBOM12724520080723

Two leading U.S. Democrats said on Tuesday they are investigating whether Food and Drug Administration knowingly allowed the sale of Ranbaxy drugs that may have been backed by fraudulent data.

The congressional investigation stems from information released as part of a probe by U.S. law enforcement authorities into whether the Indian drugmaker submitted false data to support generic drug applications and tried to hide manufacturing violations, the lawmakers said.

Reps. John Dingell and Bart Stupak said in a statement they want to know if the FDA "knowingly allowed drugs suspected of being fraudulently approved and manufactured in violation of Good Manufacturing Practices to continue being sold by Ranbaxy Inc in the United States."

Ranbaxy Laboratories Ltd, India's largest drugmaker by sales, has agreed to a $4.6 billion takeover offer from Japan's Daiichi Sankyo.

The lawmakers pointed to a legal motion filed July 3 by the Justice Department and the U.S. Attorney's Office on the FDA's behalf.

The motion, which was seeking to enforce subpoenas for documents, stated that the FDA was aware of allegations of fraudulent conduct by Ranbaxy for at least 18 months but did nothing to remove suspect products from the market, the lawmakers said.

"If true, these statements would call into serious question whether the leadership of the agency ... have met even the minimum requirements of due diligence," the lawmakers said.

FDA spokesman Christopher Kelly said the agency had received the lawmakers' letter and would respond directly to them.

Ranbaxy has denied allegations that it sold misbranded or adulterated drugs and said it was cooperating with the three-year-old probe by U.S. authorities.

Dingell chairs the House of Representatives Energy and Commerce Committee and Stupak heads its investigations subcommittee. Both are from Michigan.

The lawmakers, in a letter to the FDA, asked for documents relating to its inspections of Ranbaxy facilities and its suppliers, plus other information.

Title: **Internet helps doctor get back to basics**

Publish Date: **2008-01-28**

http://feeds.reuters.com/~r/reuters/healthNews/~3/224887128/idUSN2830900320080129

Dr. Howard Stark's office is quiet. Very quiet. No patients sit in his waiting room. No receptionist answers the telephone. Stark does not have a receptionist.

Instead, he and his assistant Michele Norris-Bell check e-mail alerts on handheld devices and -- between seeing patients in person -- on a desktop computer.

Stark has moved most of his practice, based in Washington, onto the Internet and he couldn't be happier. Since he started his Web-based service two years ago, he has received 14,000 e-mails.

And yet, he feels more like an old-fashioned family doctor in a small town than a modern, harried physician.

"That's 14,000 phone calls that we did not have to answer and that patients did not have to make," Stark said.

He does not charge for answering an e-mail. "You have to come in one time a year for an annual exam," Stark said.

The rest is free -- prescription refills, quick questions about medication, even questions about unusual stings.

"What do I get? A picture of the scorpion that bit the patient in Belize," Stark laughed. "I said, 'it would have been better to send me a picture of your leg.'"

He also gets updates on patients' personal lives.

"People say how impersonal e-mail is. No way. It is so personal because I can hear what is going on with the kids," Stark said in an interview at his otherwise ordinary office.

"It keeps me a lot closer to what is going on with my patients," he added. "I feel like I have taken 21st century medicine back to being more like the old-fashioned physician who knows how your family is doing."

Health experts, the U.S. government, labor unions, employers and average citizens all agree the U.S. health care system badly needs improvement.

SOARING COSTS, LONG WAITS

Costs are soaring and yet the average physician, according to many estimates, spends only about 10 minutes with each patient.

Harried desk staff often double- and even triple-book each appointment slot to make optimal use of the doctor's time and to make sure the overheads are covered.

Title: **Obesity seen protective in cases of heart failure**

Publish Date: **2008-08-05**

http://feeds.reuters.com/~r/reuters/healthNews/~3/356344625/idUSCOL54751920080805

Overweight and obese patients with heart failure seem to have a lower risk of dying than their normal-weight counterparts, according to a review of published studies involving more than 28,000 heart failure patients who were followed for an average of nearly three years.

There is evidence, Dr. Antigone Oreopoulos told Reuters Health, that a normal body mass index (BMI) "is likely not the ideal BMI" in people with heart failure, which occurs when the heart loses its ability to pump blood efficiently.

Oreopoulos, from University of Alberta, Edmonton, Canada, and associates reviewed nine studies that examined the impact of BMI on mortality. They pooled the data to estimate the risk of death in patients who are underweight, overweight or obese compared to patients with a normal body weight.

According to the researchers, patients who were overweight or obese were less likely to die during follow up compared to their normal-weight peers. Being overweight or obese "remained protective" against death in a "risk-adjusted" analysis.

Heart failure patients who had a normal weight or who were underweight had the highest death rates. "It remains unknown, however, if higher body fat levels are actually the cause of better outcomes in patients with heart failure," the researchers note in the American Heart Journal.

"We believe there is a need for prospective studies to confirm these findings and elucidate potential mechanisms" for the potentially protective effect of increased body weight on heart failure, Oreopoulos and colleagues conclude.

"Our findings," they point out, "are consistent with evidence in other chronic disease populations," including survivors of heart attack and chronic hemodialysis patients, demonstrating lower death rates with higher BMI levels.

Title: **Workouts boost function of insulin-making cells**

Publish Date: **2008-03-07**

http://feeds.reuters.com/~r/reuters/healthNews/~3/247623007/idUSKUA78055620080307

Sedentary older people at risk of developing diabetes showed significant improvement in the function of their insulin-making beta cells after just one week of exercise, University of Michigan researchers found.

Beta cell function is known to decline with age, although it is not clear why, Drs. Cathie J. Bloem and Annette M. Chang explain in the Journal of Clinical Endocrinology and Metabolism.

As people age, they may also become less sensitive to the blood sugar-regulating effects of insulin and develop impaired insulin secretion, the researchers add. And while short-term exercise boosts insulin sensitivity, it has not been clear how it might affect beta cell function.

To investigate, Bloem and Chang had 12 sedentary individuals aged 60 and older perform an hour-long workout every day for a week. The exercise sessions, consisting of stints on a treadmill, exercise bike and cross-training machine, required study participants to work out at 60 percent to 70 percent of their maximum heart rate capacity.

After the exercise period, study participants' sensitivity to insulin had increased by 53 percent, on average, while a measure of beta cell function called the disposition index had risen by 28 percent. However there were no changes in their fat mass, levels of fat in the blood, or other factors that might explain the effect of exercise on beta cells.

"Longer-term exercise training studies are required and are currently in progress to evaluate further exercise training effects on beta cell function in age-related glucose intolerance," the researchers note.

Title: **Red yeast rice, fish oil fight high cholesterol**

Publish Date: **2008-07-17**

http://feeds.reuters.com/~r/reuters/healthNews/~3/338291832/idUSCOL76707420080717

A regimen of supplements and lifestyle coaching is just as effective as statin medication for reducing levels of low-density lipoprotein (LDL) or "bad" cholesterol, and more effective in helping people lose weight, new research shows.

People with high cholesterol who took red yeast rice and fish oil daily and received counseling on diet, exercise and relaxation techniques showed the same 40 percent drop in LDL cholesterol seen among people taking 40 milligrams of simvastatin daily, Dr. David J. Becker of the University of Pennsylvania Health System's Chestnut Hill Hospital and colleagues found. And they pared off an average of 10 pounds over 12 weeks, compared to less than a pound for patients taking the statin.

Becker has run a lifestyle program for people at risk of heart disease for 13 years. "People had a uniform desire to get off statins, and when they did their cholesterol was only going down maybe 5 percent at most," he told Reuters Health. The cardiologist decided to launch the current study after seeing many patients have success in lowering their cholesterol with red yeast rice and fish oil.

With a grant from the state of Pennsylvania, Becker and his team randomly assigned 74 patients to receive 40 milligrams of simvastatin (Zocor) daily along with printed information on lifestyle changes, or to three capsules of fish oil twice daily and 600 milligrams of red yeast rice daily along with the 12-week lifestyle program.

LDL cholesterol levels fell by 42.4 percent in the red yeast rice group and by 39.6 percent in the simvastatin group, not a statistically significant difference. Triglyceride levels didn't change in the statin group, but fell 29 percent in the red yeast rice group, probably because they were taking fish oil, according to Becker and his team.

People in the red yeast rice group lost an average of 4.7 kilograms (just over 10 pounds), compared to 0.3 kilograms (less than a pound) in the statin group.

Red yeast rice comes from fermenting red yeast with rice. Known as hong ku, the substance has been used as a medicine and food garnish in parts of Asia for centuries, Becker said. It contains a substance called monacolin-K that is nearly identical to the cholesterol-lowering drug lovastatin (Mevacor), as well as several other monacolins that may also have cholesterol-lowering properties.

People in the red yeast rice arm of the study were taking the equivalent of 10 to 15 mg of lovastatin, Becker said. "This lovastatin dosage is quite small, yet the effects we saw with the red yeast rice were akin to those one would generally see with a much higher dose of lovastatin."

"However, it is not risk-free, and it must be used carefully and in conjunction with your physician."

If more studies bear out the current findings, he added, the supplement/lifestyle intervention he and his colleagues tested could offer an alternative to people with high cholesterol who don't want to take statins, or who can't tolerate the drugs. However, he added, people who actually have heart disease should stick with statins, because they have been shown to reduce mortality.

Becker noted that a recent analysis by ConsumerLab found red yeast rice products varied sharply in their potency, and some were contaminated with a toxic byproduct called citrinin. "This paper is a call for better regulation of this supplement as well so that we know consistently what's in it," he said.

Title: **Americans losing sleep over financial crisis**

Publish Date: **2008-10-27**

http://feeds.reuters.com/~r/reuters/healthNews/~3/EkPP4ZevIE4/idUSTRE49Q7T720081027

If fellow workers seem groggier or grumpier than usual in the mornings, they are probably losing sleep over the global financial crisis, according to research released on Monday.

Ninety-two percent of respondents said the economic turmoil is keeping them awake at night, according to a survey by ComPsych Corp, a provider of employee assistance programs.

Of those, a third said their biggest worry was the cost of living, while another third cited their credit card debt.

One in six said their biggest worry was their mortgage payment, and another one in six cited concern over their retirement account.

Eight percent of those surveyed said they were not worried.

Chicago-based ComPsych conducted the online survey of 1,137 employed adults across the United States from October 6 through October 17. The margin of error was plus or minus 3 percentage points.

Title: **Doctor group supports trans-fat bans**

Publish Date: **2008-10-27**

http://feeds.reuters.com/~r/reuters/healthNews/~3/qtTidyQHtvI/idUSTRE4A982M20081110

The American Medical Association threw its weight behind legislation to ban the use of artificial trans fats in restaurants and bakeries nationwide on Monday.

The group, which represents about 240,000 doctors and medical students, said it would move away from a previous, gentler position that advised people to reduce their use and move to healthier fats and oils instead.

"Trans fats have been proven to raise LDL (low density lipoprotein), the bad cholesterol, while lowering HDL (high density lipoprotein), the good cholesterol, which significantly increases the risk for heart disease," said AMA board member Dr. Mary Anne McCaffree.

"By supporting a ban on the use of artificial trans fats in restaurants and bakeries, we can help improve the quality of the food Americans eat and may ultimately save lives."

The group, meeting in Orlando, Florida, said replacing trans fats would prevent up to 100,000 premature deaths each year in the United States alone.

Trans fats come from adding hydrogen to vegetable oil through a process called hydrogenation. It makes liquid oil more like butter and makes it less likely to go rancid -- but in the process makes it just as dangerous to arteries as butter or lard.

New York City and California banned trans fats in July.

Title: **Increase in throat cancer parallels obesity rate**

Publish Date: **2008-04-03**

http://feeds.reuters.com/~r/reuters/healthNews/~3/263506039/idUSCOL36818320080403

The rising incidence of throat cancer, also referred to as cancer of the esophagus or esophageal adenocarcinoma, may be related to Americans' increasing intake of total and refined carbohydrates and subsequent rise in obesity rates.

"The similarity in these trends gives further evidence for the association of carbohydrate intake, obesity, and related measures with cancer," Dr. Cheryl L. Thompson told Reuters Health.

She and colleagues caution, however, that such observations do not necessarily reflect individual risk for esophageal adenocarcinoma

The researchers, all associated with Case Western Reserve University and University Hospitals of Cleveland, in Ohio, compared National Cancer Institute data for esophageal adenocarcinoma (1973-2001) and food consumption information from the National Nutrient Data Bank (1909-1997).

The incidence of esophageal adenocarcinoma increased over the review period and "strongly correlated" with carbohydrate consumption. This cancer is also known to be strongly associated with gastroesophageal reflux (GERD), which, in turn, associated with obesity and a high carbohydrate intake, the investigators report in the American Journal of Gastroenterology.

By contrast, they noted a decrease in the rates of squamous cell cancer of the esophagus, which is more closely associated with smoking rather than reflux disease and obesity.

The researchers found a trend toward higher intakes of refined carbohydrates; those with more starch and lower nutrient levels than carbohydrates obtained from whole grains and minimally processed foods.

These findings highlight the importance of limiting refined carbohydrates in the American diet, the investigators note. Additional research is needed to assess individual risk from high intake of refined carbohydrates, Thompson adds.

Title: **Child's sleep linked to adulthood obesity risk**

Publish Date: **2008-11-03**

http://feeds.reuters.com/~r/reuters/healthNews/~3/Hp8KWRVjT7Y/idUSTRE4A24DL20081103

Consistently getting a good night's sleep may help protect children from becoming obese as adults, a study published Monday suggests.

Researchers found that among more than 1,000 people followed from birth to age 32, those who got too little sleep as children were more likely than their well-rested counterparts to become obese adults.

Even with a range of other factors considered -- like childhood weight and TV habits, and adulthood exercise levels -- there remained a link between sleep deprivation during childhood and obesity risk later in life.

All of this supports the idea that early sleep habits have a direct effect on weight in the long term, according to Dr. Robert John Hancox, the study's senior author.

"Although we cannot prove that this is a cause-and-effect relationship," he told Reuters Health, "this study provides strong evidence that it probably is."

Hancox and his colleagues at the University of Otago in Dunedin, New Zealand, report the findings in the journal Pediatrics.

A number of studies have found that sleep-deprived adults and children are at greater risk of being overweight. However, this is the first study to show a long-term relationship between sleep and obesity risk, Hancox said.

The study involved 1,037 men and women who had been followed since their birth, between 1972 and 1973, up to the age of 32. When the participants were 5, 7, 9 and 11 years old, their parents reported on their usual bed time and wake-up time.

In general, Hancox and his colleagues found, as childhood sleep time declined, adulthood body mass index, or BMI, climbed.

Adults who had been "short sleepers" as children -- averaging fewer than 11 hours in bed each night -- generally had a higher BMI than those who'd gotten more sleep as kids.

"Importantly, this is not because children who were short sleepers grew up to be short sleepers as adults," Hancox pointed out. "In other words, inadequate sleep in childhood appears to have long-lasting consequences."

The findings, according to the researchers, suggest that weight control may stand as another reason for children to get a good night's sleep. Experts generally recommend that children between the ages of 5 and 12 sleep for about 11 hours each night, while teenagers should get 8.5 to 9.5 hours.

It's thought that children today are getting less sleep than the generations before them did, Hancox noted. That trend, he added, could be helping to feed the rise in obesity.

No one knows for certain why lack of sleep is linked to heavier weight. One theory, based on research in the sleep lab, is that sleep deprivation alters the normal balance of appetite-stimulating and appetite-suppressing hormones. Sleepy children may also be too tired for physical activity during the day.

Title: **Studies show exercise boon for obesity, diabetes**

Publish Date: **2008-7-28**

http://feeds.reuters.com/~r/reuters/healthNews/~3/348761444/idUSTON87483720080728

Walking a bit more each day can help people control their Type 2 diabetes but obese people trying to keep weight off may need to exercise harder than they had thought, according to a studies published on Monday.

Simply walking 45 minutes more each day helped people with diabetes use blood sugar better, Michael Trenell of Britain's Newcastle University and colleagues wrote in the journal Diabetes Care.

"People often find the thought of going to the gym quite daunting, but what we've found is that nearly everyone with diabetes is able to become more active through walking," Trenell said.

The Newcastle team paired 10 Type 2 diabetes patients with people without the condition of similar height, weight and age and asked everybody to walk more than 10,000 steps each day.

Magnetic resonance imaging or MRI scans showed that people who walked 45 minutes more each day burned about 20 percent more fat -- increasing the ability of the muscles to store sugar and help control diabetes, the researchers said.

"What is exciting about this study is that it provides an immediate way to help control diabetes without any additional drugs," Trenell said.

Diabetes affects an estimated 246 million adults worldwide and accounts for 6 percent of all global deaths. Type 2 diabetes accounts for about 90 percent of all diabetes cases and is closely linked to obesity and physical inactivity.

Obesity and diabetes both are growing problems as more developing nations adopt a Western lifestyle, something the International Diabetes Federation estimates will propel the number of people with diabetes to 380 million by 2025.

But current exercise guidelines calling for people to get 150 minutes -- 2.5 hours -- each week may not be enough to help the obese keep weight off, John Jakicic of the University of Pittsburgh and colleagues wrote in the Archives of Internal Medicine.

To determine an optimal amount of exercise, the U.S. team enrolled 201 overweight and obese women in a weight loss programme between 1999 and 2003 and assigned them to one of four exercise groups.

After six months, women in all four groups had lost an average of 8 to 10 percent of their weight but many gained it back.

Women assigned to exercise for about an extra hour each day did not gain the weight back, the researchers said. These women were also more likely to stick to healthy diets.

Jakicic recommended that people who want to lose weight and keep it off get at least 4-1/2 hours of exercise a week.

"There is a growing consensus that more exercise may be necessary to enhance long-term weight loss," Jakicic and colleagues wrote.

APPENDIX E. SalientHealthNews POPULATION HEALTH NEWS DATA

Gold standard Lifestyle framing feature score averages

Feature	Average Life framing (+)	Average Life framing (-)	*p*-value	AUC
length	2019.0	1701.0	**0.005**	**0.678**
activation	1.371	1.319	**0.006**	**0.669**
imagery	1.275	1.211	**0.001**	**0.703**
pleasantness	1.463	1.402	**0.002**	**0.733**
friend	0.0030	0.0020	0.080	0.597
humans	0.0233	0.0169	**0.025**	**0.632**
cause	0.0172	0.0216	0.053	**0.64**
certain	0.0056	0.0040	0.077	0.598
bio	0.0984	0.0863	0.093	**0.631**
body	0.0199	0.0197	0.955	0.599
health	0.0333	0.0200	**0.013**	0.516
ingest	0.0165	0.0053	**0.001**	0.556
leisure	0.0023	0.0008	0.160	**0.661**

Gold standard biological/genetic causation framing feature score averages

Feature	Average Bio Framed (+)	Average Bio Framed (-)	p	AUC
length	1867.6	1852.4	0.895	0.581
activation	1.331	1.359	0.152	0.581
imagery	1.233	1.252	0.356	0.518
pleasantness	1.416	1.449	0.093	0.564
friend	0.0028	0.0021	0.184	0.579
humans	0.0227	0.0176	0.077	0.593
cause	0.0206	0.0182	0.292	**0.626**
certain	0.0046	0.0050	0.744	**0.712**
bio	0.1023	0.0824	**0.005**	**0.603**
body	0.0265	0.0131	**<0.001**	**0.648**
health	0.0284	0.0250	0.527	**0.661**
ingest	0.0082	0.0136	0.133	0.576
leisure	0.0010	0.0021	0.302	**0.61**

Obesity news

Feature	Average score (+)	p vs lifestyle	p vs bio
length	1869.0	(⇓) 0.109	
activation	1.337	(⇓) 0.023	
imagery	1.233	(⇓) 0.003	
pleasantness	1.428	(⇓) 0.030	
friend	0.0029	(⇓) 0.790	
humans	0.0210	(⇓) 0.288	(⇑) 0.403
cause	0.0194	(⇑) 0.134	(⇓) 0.420
certain	0.0047	(⇓) 0.205	
bio	0.0985		(⇑) 0.403
body	0.0203		(⇑) 0.060
health	0.0294		(⇑) 0.787
ingest	0.0097	(⇓) 0.034	(⇑) 0.548
leisure	0.0014	(⇓) 0.419	

African-Americans/Blacks health news

	News discussing African-Americans/Blacks across all health news			**News discussing African-Americans/Blacks across obesity news**		
Feature	mean	Lifestyle framing (trend) p	Bio/genetic framing (trend) p	mean	Lifestyle framing (trend) p	Bio/genetic framing (trend) p
length	1804.2	(⇓) 0.044		2003.2	(⇓) 0.936	
activation	1.349	(⇓) 0.195		1.356	(⇓) 0.624	
imagery	1.231	(⇓) 0.010		1.237	(⇓) 0.201	
pleasantness	1.433	(⇓) 0.122		1.446	(⇓) 0.572	
friend	0.0019	(⇓) 0.020		0.002	(⇓) 0.152	
humans	0.0175	(⇓) 0.043	(⇑) 0.058	0.0212	(⇓) 0.645	(⇓) 0.747
cause	0.0163	(⇓) 0.647	(⇓) 0.034	0.0182	(⇑) 0.819	(⇓) 0.604
certain	0.0054	(⇓) 0.832		0.0077	(⇑) 0.530	
bio	0.0783		(⇓) <0.001	0.0995		(⇓) 0.847
body	0.0155		(⇓) 0.004	0.0191		(⇓) 0.411
health	0.0058		(⇓) <0.001	0.0227		(⇓) 0.451
ingest	0.005	(⇓) <0.001	(⇓) 0.186	0.0052	(⇓) 0.007	(⇓) 0.375
leisure	0.0012	(⇓) 0.323		0	(⇓) 0.033	

Elderly/Seniors health news

	News discussing Elderly/Seniors across all health news			News discussing Elderly/Seniors across obesity news		
Feature	mean	Lifestyle framing (trend) p	Bio/genetic framing (trend) p	mean	Lifestyle framing (trend) p	Bio/genetic framing (trend) p
length	1792.1	(⇓) 0.023		1890.4	(⇓) 0.338	
activation	1.336	(⇓) 0.029		1.352	(⇓) 0.483	
imagery	1.24	(⇓) 0.023		1.275	(⇑) 0.988	
pleasantnes s	1.422	(⇓) 0.020		1.442	(⇓) 0.511	
friend	0.0026	(⇓) 0.454		0.003	(⇑) 0.970	
humans	0.0192	(⇓) 0.100	(⇑) 0.138	0.0197	(⇓) 0.332	(⇓) 0.412
cause	0.0189	(⇑) 0.296	(⇓) 0.329	0.0204	(⇑) 0.288	(⇓) 0.945
certain	0.0046	(⇓) 0.219		0.004	(⇓) 0.245	
bio	0.0737		(⇓) <0.001	0.0984		(⇓) 0.701
body	0.0175		(⇓) 0.015	0.0291		(⇑) 0.674
health	0.0041		(⇓) <0.001	0.0132		(⇓) 0.006
ingest	0.0076	(⇓) 0.009	(⇑) 0.799	0.0112	(⇓) 0.314	(⇑) 0.538
leisure	0.0023	(⇓) 0.966		0.0005	(⇓) 0.110	

Child/Adolescent health news

	News discussing Children/Adolescents across all health news			News discussing Children/Adolescents across obesity news		
Feature	mean	Lifestyle framing (trend) p	Bio/genetic framing (trend) p	mean	Lifestyle framing (trend) p	Bio/genetic framing (trend) p
length	1679.7	(⇓) <0.001		1913.8	(⇓) 0.285	
activation	1.318	(⇓) <0.001		1.342	(⇓) 0.068	
imagery	1.229	(⇓) 0.001		1.235	(⇓) 0.010	
pleasantness	1.415	(⇓) 0.003		1.449	(⇓) 0.413	
friend	0.0021	(⇓) 0.032		0.0029	(⇓) 0.849	
humans	0.0269	(⇑) 0.104	(⇑) 0.036	0.0294	(⇑) 0.012	(⇑) 0.003
cause	0.021	(⇑) 0.010	(⇓) 0.798	0.0181	(⇑) 0.553	(⇓) 0.135
certain	0.0054	(⇓) 0.721		0.0045	(⇓) 0.156	
bio	0.0717		(⇓) <0.001	0.0852		(⇓) 0.002
body	0.0117		(⇓) <0.001	0.0154		(⇓) 0.002
health	0.0145		(⇓) <0.001	0.0308		(⇑) 0.559
ingest	0.0076	(⇓) 0.006	(⇑) 0.772	0.011	(⇓) 0.102	(⇑) 0.291
leisure	0.0026	(⇑) 0.776		0.0028	(⇑) 0.718	

Hispanic/Latino health news

	News discussing Hispanics/Latinos across all health news			**News discussing Hispanics/Latinos across obesity news**		
Feature	mean	Lifestyle framing (trend) p	Bio/genetic framing (trend) p	mean	Lifestyle framing (trend) p	Bio/genetic framing (trend) p
length	1755.8	(⇓) 0.020		1876.7	(⇓) 0.322	
activation	1.336	(⇓) 0.079		1.388	(⇑) 0.468	
imagery	1.221	(⇓) 0.007		1.283	(⇑) 0.773	
pleasantness	1.425	(⇓) 0.118		1.518	(⇑) 0.043	
friend	0.0026	(⇓) 0.736		0.0055	(⇑) 0.509	
humans	0.0199	(⇓) 0.348	(⇑) 0.435	0.0315	(⇑) 0.088	(⇑) 0.067
cause	0.0128	(⇓) 0.035	(⇓) <0.001	0.0143	(⇓) 0.446	(⇓) 0.120
certain	0.0062	(⇑) 0.661		0.0112	(⇑) 0.057	
bio	0.0744		(⇓) <0.001	0.0742		(⇓) 0.051
body	0.0127		(⇓) 0.004	0.0084		(⇓) 0.021
health	0.0101		(⇓) 0.001	0.03		(⇑) 0.880
ingest	0.0062	(⇓) 0.004	(⇑) 0.466	0.0119	(⇓) 0.426	(⇑) 0.503
leisure	0.0023	(=) 1.000		0.0047	(⇑) 0.544	

Gay/MSM/homosexual/bisexual (GMHB) health news

	News discussing GMHB across all health news		
Feature	mean	Lifestyle framing (trend) p	Bio/genetic framing (trend) p
length	1670.6	(⇓) 0.033	
activation	1.312	(⇓) 0.029	
imagery	1.22	(⇓) 0.020	
pleasantness	1.389	(⇓) 0.006	
friend	0.0011	(⇓) 0.005	
humans	0.0319	(⇑) 0.061	(⇑) 0.042
cause	0.016	(⇓) 0.547	(⇓) 0.031
certain	0.0079	(⇑) 0.269	
bio	0.0767		(⇓) 0.007
body	0.0018		(⇓) <0.001
health	0.0015		(⇓) <0.001
ingest	0.0068	(⇓) 0.013	(⇑) 0.650
leisure	0.0011	(⇓) 0.295	

Men's health news

	News discussing men across all health news			News discussing men across obesity news		
Feature	mean	Lifestyle framing (trend) p	Bio/genetic framing (trend) p	mean	Lifestyle framing (trend) p	Bio/genetic framing (trend) p
length	1721.3	(⇓) 0.002		1900	(⇓) 0.238	
activation	1.338	(⇓) 0.026		1.347	(⇓) 0.169	
imagery	1.244	(⇓) 0.026		1.25	(⇓) 0.149	
pleasantness	1.426	(⇓) 0.021		1.442	(⇓) 0.250	
friend	0.0026	(⇓) 0.443		0.0033	(⇑) 0.479	
humans	0.03	(⇑) 0.004	(⇑) <0.001	0.0317	(⇑) 0.001	(⇑) <0.001
cause	0.0185	(⇑) 0.384	(⇓) 0.172	0.018	(⇑) 0.659	(⇓) 0.150
certain	0.0048	(⇓) 0.222		0.0039	(⇓) 0.032	
bio	0.0879		(⇑) 0.006	0.1008		(⇓) 0.811
body	0.0198		(⇑) 0.044	0.0256		(⇓) 0.808
health	0.0128		(⇓) <0.001	0.0298		(⇑) 0.763
ingest	0.0065	(⇓) 0.003	(⇑) 0.473	0.0076	(⇓) 0.011	(⇓) 0.816
leisure	0.0016	(⇓) 0.505		0.0014	(⇓) 0.441	

Low income/socio-economic (SES) status health news

	News discussing low SES across all health news			News discussing low SES across obesity news		
Feature	mean	Lifestyle framing (trend) p	Bio/genetic framing (trend) p	mean	Lifestyle framing (trend) p	Bio/genetic framing (trend) p
length	1776.2	(⇓) 0.050		2105	(⇑) 0.844	
activation	1.35	(⇓) 0.268		1.342	(⇓) 0.655	
imagery	1.255	(⇓) 0.334		1.249	(⇓) 0.782	
pleasantness	1.444	(⇓) 0.361		1.448	(⇓) 0.854	
friend	0.0029	(⇓) 0.978		0.0027	(⇓) 0.604	
humans	0.0208	(⇓) 0.446	(⇑) 0.557	0.0241	(⇑) 0.912	(⇑) 0.838
cause	0.014	(⇓) 0.077	(⇓) <0.001	0.0158	(⇓) 0.819	(⇓) 0.466
certain	0.0049	(⇓) 0.557		0	(⇓) <0.001	
bio	0.0618		(⇓) <0.001	0.1006		(⇓) 0.912
body	0.0082		(⇓) <0.001	0.0098		(⇓) <0.001
health	0.008		(⇓) <0.001	0.0324		(⇑) 0.690
ingest	0.0042	(⇓) <0.001	(⇓) 0.104	0.0112	(⇓) 0.403	(⇑) 0.613
leisure	0.0052	(⇑) 0.078		0.0054	(⇑) 0.627	

Women's health news

	News discussing women across all health news			News discussing women across obesity news		
Feature	mean	Lifestyle framing (trend) p	Bio/genetic framing (trend) p	mean	Lifestyle framing (trend) p	Bio/genetic framing (trend) p
length	1714.2	(⇓) 0.002		1915.1	(⇓) 0.292	
activation	1.336	(⇓) 0.016		1.346	(⇓) 0.107	
imagery	1.237	(⇓) 0.007		1.237	(⇓) 0.014	
pleasantness	1.428	(⇓) 0.029		1.448	(⇓) 0.361	
friend	0.0026	(⇓) 0.379		0.0034	(⇑) 0.347	
humans	0.0283	(⇑) 0.024	(⇑) 0.006	0.0318	(⇑) <0.001	(⇑) <0.001
cause	0.0179	(⇑) 0.607	(⇓) 0.078	0.0169	(⇓) 0.841	(⇓) 0.029
certain	0.0051	(⇓) 0.476		0.0049	(⇓) 0.380	
bio	0.0847		(⇓) <0.001	0.1004		(⇓) 0.740
body	0.0179		(⇓) 0.009	0.0213		(⇓) 0.143
health	0.0127		(⇓) <0.001	0.0294		(⇑) 0.815
ingest	0.007	(⇓) 0.004	(⇑) 0.603	0.0087	(⇓) 0.020	(⇑) 0.850
leisure	0.0015	(⇓) 0.462		0.0014	(⇓) 0.404	

Rural population health news

	News discussing rural populations across all health news			News discussing rural populations across obesity news		
Feature	mean	Lifestyle framing (trend) p	Bio/genetic framing (trend) p	mean	Lifestyle framing (trend) p	Bio/genetic framing (trend) p
length	1756.3	(⇓) 0.178		2272.5	(⇑) 0.236	
activation	1.319	(⇓) 0.017		1.361	(⇓) 0.525	
imagery	1.25	(⇓) 0.120		1.254	(⇓) 0.148	
pleasantness	1.421	(⇓) 0.084		1.49	(⇑) 0.135	
friend	0.0014	(⇓) 0.015		0.0013	(⇓) 0.398	
humans	0.0152	(⇓) 0.035	(⇓) 0.046	0.0206	(⇓) 0.830	(⇓) 0.870
cause	0.0177	(⇑) 0.840	(⇓) 0.234	0.0143	(⇓) 0.551	(⇓) 0.286
certain	0.0041	(⇓) 0.241		0.0051	(⇓) 0.931	
bio	0.0655		(⇓) 0.002	0.0643		(⇓) 0.186
body	0.0076		(⇓) <0.001	0.0043		(⇓) 0.042
health	0.0073		(⇓) 0.001	0.0336		(⇑) 0.325
ingest	0.0098	(⇓) 0.065	(⇑) 0.612	0.0094	(⇓) 0.030	(⇑) 0.635
leisure	0.0032	(⇑) 0.673		0	(⇓) 0.033	

Health news discussing populations with health disparities (African-American/Hispanics/Low SES/Rural)

	News discussing populations with health disparities across all health news			News discussing populations with health disparities across obesity news		
Feature	mean	Lifestyle framing (trend) p	Bio/genetic framing (trend) p	mean	Lifestyle framing (trend) p	Bio/genetic framing (trend) p
length	1776.2	(⇓) 0.017		2021.1	(⇑) 0.988	
activation	1.346	(⇓) 0.110		1.364	(⇓) 0.739	
imagery	1.242	(⇓) 0.027		1.259	(⇓) 0.486	
pleasantness	1.436	(⇓) 0.113		1.475	(⇑) 0.633	
friend	0.0023	(⇓) 0.199		0.0033	(⇑) 0.804	
humans	0.0194	(⇓) 0.120	(⇑) 0.165	0.0248	(⇑) 0.679	(⇑) 0.533
cause	0.0155	(⇓) 0.287	(⇓) 0.003	0.0162	(⇓) 0.718	(⇓) 0.119
certain	0.0056	(⇓) 0.965		0.0076	(⇑) 0.337	
bio	0.0725		(⇓) <0.001	0.0893		(⇓) 0.167
body	0.0131		(⇓) <0.001	0.0136		(⇓) 0.024
health	0.007		(⇓) <0.001	0.0279		(⇓) 0.937
ingest	0.0054	(⇓) <0.001	(⇑) 0.230	0.0083	(⇓) 0.029	(⇑) 0.988
leisure	0.0027	(⇑) 0.774		0.0025	(⇑) 0.935	