

# **Multiplicity and Uncertainty: Media Coverage of Autism Causation**

## **Yujia Zhai**

*Department of Information Resource Management, School of Management, Tianjin Normal University, Tianjin 300387, China.*

*Telephone: (86-22) 2376-6058*

*E-mail: [zhaiyujiachn@gmail.com](mailto:zhaiyujiachn@gmail.com)*

## **Shaojing Sun (Corresponding Author)**

*School of Journalism, Fudan University, Shanghai 200433, China.*

*400 Guoding Rd, WuJiaoChang, Yangpu Qu, Shanghai Shi, China, 200433*

*Telephone: (86-21)6564-3695*

*E-mail: [shaojingsun@gmail.com](mailto:shaojingsun@gmail.com)*

## **Fang Wang**

*Department of Information Resource Management, Business School, Nankai University, Tianjin 300071, China.*

*Telephone: (86-22) 2350-2800*

*E-mail: [wangfangnk@nankai.edu.cn](mailto:wangfangnk@nankai.edu.cn)*

## **Ying Ding**

*School of Informatics & Computing, Indiana University, Bloomington, IN, 47405, USA;*

*School of Information Management, Wuhan University, Wuhan, Hubei 430072, China;*

*University Library, Tongji University, Shanghai 200092, China.*

*Telephone: (812) 855-5388*

*E-mail: [dingying@indiana.edu](mailto:dingying@indiana.edu)*

# Multiplicity and Uncertainty: Media Coverage of Autism Causation

## Abstract

Employing the machine learning method, this study analyzes 6,504 articles from four major newspapers, *New York Times*, *Washington Post*, *USA Today*, and *The Guardian*, to examine how media cover the topic about causes of autism. A total of 14,305 causal sentences on the topic are extracted from media articles and subjected to analysis of causal entities and descriptions. Results show media have presented multiple factors (e.g. vaccination, genetics, and parenting) pertaining to the causes of autism, as well as multiple symptoms of autism. Most of those causal relationships are presented in a tentative or uncertain manner. The study also reveals significant differences in reportage of autism causation across time and media channels.

**Keywords:** Autism Causation; Media Coverage; Machine Reading

## 1 Introduction

Over the past decades, the field of autism spectrum disorders (ASDs) has been expanding and the volume of relevant scientific publications has been growing exponentially (Matson & LoVullo, 2009). Among the broad category of ASDs, autism is most intensely studied, followed by Rett Syndrome and Asperger's Syndrome. Meanwhile, finding the causes and potential cures for autism has become a spotlight topic for media coverage (McKeever, 2012). The growing media attention to the ASD is echoed by scientific statistics reporting that autistic cases have increased by 20-fold over the last two decades (Diagnostic and Statistical Manual of Mental Disorders, DSM-5).

Although using experts' judgments or meta-analyses of academic publications may help locate potential causes of ASD too, inspecting newspaper coverage on the causes of autism is of unique and tremendous values. Concretely, popular media remain to be a pivotal information source that parents rely on in making a decision on the treatment for autism (Miller, Schreck, Mulick, & Butter, 2012). Recently, Schreck, Russell, and Vargas (2013) reported that more than 75% treatments of autism mentioned in print media (e.g., newspapers and magazines) are not scientifically supported. Examining newspaper coverage, as such, will provide insight into information which may significantly affect parents' decision on treatment for autism.

The mutual relationship and influence between scientific research and media coverage, however, could be complicated and consequential. For instance, Phillips, Kanter, Bednarczyk, and Tastad (1991) found the coverage of medical research in the *New York Times* significantly boosted the visibility of academic research, which subsequently received much higher attention in terms of citations from the scientific community. The authors further admonished that popular media coverage of scientific research could be enhancing or distorting the transmission process of information

flowing from academia to other communities.

In light of the longtime controversy around autism, it is particularly meaningful to examine how media cover a health issue of the like. Such a research endeavor will enhance the understanding of how media communication may present, either in a neutral or biased manner, scientific research on autism. A cursory review of existing literature reveals that little research had been undertaken to investigate media coverage of the etiology and consequences of autism, and existing research on the topic is typically limited to a small sample of media articles (Clarke, 2008). The present study therefore focuses on how media cover different causes of autism, and furthermore how such coverage may vary across time and media sources.

Employing the method of machine reading (e.g., OpenIE), we explore the patterns of reportage on the etiology of autism in a more comprehensive manner. With a dataset much larger than those used in prior media content analysis (e.g., Clarke, 2008), our study may provide insight into the cause-and-effect issues of autism as portrayed in media. The study may thus serve as a stepping stone for bridging the gap of understandings on ADS between the scientific community, media professionals, and the general public. More importantly, as an interdisciplinary study of information science and communication, the adoption of machine reading for knowledge mining provides a new perspective for understanding information dissemination and cognitive construction.

We begin with a review of relevant literature on ADS causation and media coverage, including linkage to autism, and the development of machine reading. We then introduce our research questions and explain the machine-learning method employed in the current study. We present our results in a textual and visual manner, and finally, interpret the implications of our research findings in reference to research literature and media practice. We conclude the study with research limitations and future directions.

## **2 Related Work**

### ***2.1 Causations and Consequences of Autism***

The etiology, causation, and consequences of autism have long been contested and controversial (Currenti, 2010). As Newschaffer et al. (2007) commented, the causal mechanisms underlying autism have not been well studied or understood. As a result, different groups or communities of people may hold varying and even conflicting views on the causes and consequences of autism. For instance, Mercer, Creighton, Holden, and Lewis (2006) surveyed 41 parents with autistic children, and summarized the parental perspectives on causes of autism, where reported causes of autism were ranked as: genetic influences (90.2%), perinatal factors (68.3%), diet (51.2%), prenatal factors (43.9%), and vaccines (40.0%).

In the scientific community, researchers have also explored a variety of factors in autism causation, such as socio-economic background, and genetic and environmental

influences. McKeever (2012), after content-analyzing the New York Times and the Washington Post, found that science frames (i.e. discussing the genetic, biological, and environmental causes and effects of autism) were prevalent in the early years of news coverage of autism. The prevalence of scientific frames has remained largely consistent, though it shows a slight decrease across time. The authors further ascribed this prevalence to the fact that scientists have not identified the exact cause or causes of autism. In addressing the history of “autism blame” (PBS, 2002), for example, earlier conceptualizations of autism causes featured psychogenic factors such as emotional distance or coldness of parents (notably “refrigerator mothers” and absent fathers). Those factors contribute to a distinct neurological disorder identified by Leo Kanner in 1943 as “Early Infantile Autism,” which he incorrectly associated with high-income, white families.

An important long-term task is to explore and understand how genetic and environmental factors contribute to autism (Currenti, 2010). Researchers have examined a wide range of genes for potential associations with the condition, though, none has been replicated or confirmed as a definitive autism gene (Newschaffer et al., 2007). Similarly, the linkage between autism and environmental factors has attracted increasing research attention, especially in light of increased environmental toxins. Yet little evidence has been accumulated to confirm that linkage (Currenti, 2010).

Factors compounding the understanding and conceptualization of the condition are also diverse and complex. For instance, Gernsbacher, Dawson, Goldsmith (2005) pointed out that the changing diagnostic criteria, along with other factors, are leading to further misunderstandings or disagreements of autism’s causes. Verhoeff (2013a) reviewed historical changes in autism and divided the period into three stages: affective aloofness and withdrawal, language and cognitive abnormalities, and deficits in intuition and social cognition. The author contended, however, that the definition of autism is not fixed but rather is constantly in flux. The various conceptualizations and diagnostic criteria, among other factors, exert significant influences on the quality and conclusion of different studies. As a result, even scientific researchers are not able to reach a consensus on the causation and effects of autism. The controversial nature of the conditions, causes, and related issues also complicates and diversifies how ADS is handled by the media.

## ***2.2 Media Coverage of Autism***

Clarke (2008) elucidated the two approaches to media coverage of controversial health/science issues, structural and functional. The structural approach focuses on scrutinizing the actual messages conveyed by media, whereas the functional approach focuses on examining the effects those media messages may give rise to. For the present study, we believe inspecting the messages on the media best serves as a precursory step to understanding the effects of those messages on public audiences.

One topic receiving intense media coverage is the potential use of vaccinations in causing autism. The autism-vaccine controversy can be dated back to the publication of a well-publicized 1998 study published by Wakefield et al. in *Lancet*. That essay is

retracted by *Lancet* after being discredited and in light of what is generally seen as an overwhelming body of research evidence has negated the link between autism and vaccine usage. But many people, including parents whose children reportedly experienced extreme reactions to certain vaccines, are not convinced that this connection does not exist (WebMD Special Report, 2016). Some of those concerns have been fully or partially testified. For example, federal authorities stated that a George girl could receive compensation because vaccines may have “aggravated an underlying condition, causing autism-like symptoms,” and NIH confirmed that they are looking very carefully into numerous parents’ reports of autism symptoms following vaccines (Downs, 2016). Clarke (2008) analyzed 279 news articles on the autism-vaccine controversy, and found that 31 percent of the articles cited both pro-link and anti-link studies/claims. In more than 40 percent of the articles, however, only pro-link or anti-link claims were presented.

Media exercise a significant impact on public understanding of and social action on a range of health issues. Brown, Zavestoski, McCormick, Mandelbaum, and Luebke (2001) found that print media do not pay much attention to environmental causes of breast cancer, but rather provide more coverage of dietary factors or genetic causes of the disease. The authors criticized the lack of media coverage of environmental causation of breast cancer, and noted this could impede the scientific and activist combat of environmental contaminants bearing on this increasingly common disease. Brown et al. (2001) found that media reportage on environmental causation of breast cancer was presented in an uncertain way, in that media coverage relies heavily on rhetorical questions or qualifying words/sentences (e.g. there “might” be a link, “possibly” or “wonder”) that convey uncertainty. Parrott et al. (2004) also pointed out that one’s personal belief about the cause(s) of a particular illness may greatly affect one’s knowledge of and response to the disease.

### **2.3 Machine Reading**

Machine reading (MR) contributes to the automatic, non-human supervised understanding of texts, with a focus on processing large text collections and extracting knowledge bases of concrete facts (Cronin, van Leeuwen, Mulenga, & Bodamer, 2011; Etzioni, 2007). Knowledge facts in this system are usually sentences broken down and encoded in the form of subject-predicate-object tuples. One major task in MR is Open Information Extraction (OIE), which automatically extracts fact tuples from large, heterogeneous, and unstructured texts. It treats arbitrary noun phrases as subjects or objects and arbitrary verb phrases as predicates. MR can garner direct knowledge mapping between the text and the ontology to draw conclusions about implicitly given knowledge. An MR system relies largely on Machine Learning and Natural Language Process (NLP) technologies, such as parsing, role labeling, entity recognition, anaphora resolution, relation extraction (Poon & Domingos, 2010) and cross-document co-reference resolution.

More recent work on machine reading has focused on building large-scale knowledge bases such as Reverb, WOE parse, Ollie, and DeepDive (Fader, Soderland, & Etzioni, 2011; Ghosh, Shankar, & Owre, 2011; Poon & Vanderwende, 2010;

Schmitz, Bart, Soderland, & Etzioni, 2012; Weld & Wu, 2010). It also targets big industry players such as Google (Singhal, 2012) and IBM (Ferrucci, 2012), which are moving fast towards large-scale knowledge-oriented systems. FAUST (Flexible Acquisition and Understanding System for Text) is a machine-reading system for extracting specific information from natural language texts such as events from news articles. For example, given a sport news paragraph with different terms and scores, FAUST can automatically identify the game winner and loser (Ghosh et al., 2011). Importantly, MR is used to accelerate medical breakthroughs such as helping scientists find new cures for a range of diseases. Similar to this study, Tsutsui, Ding and Meng (2016) applied a machine-reading approach to better gauge and understand the overwhelming literature on the causes and conditions related to Alzheimer's disease.

Machine reading combines OIE and making inferences from the original extracted tuples, which are finite as ordered lists of elements. In our work, a tuple is consist of argument phrases from the sentence and a phrase that expresses a relation between the arguments, in the format (arg1; rel; arg2). There are several existing systems for OIE tasks. (1) REVERB (Fader et al., 2011), which uses shallow syntactic processing to identify relation phrases that begin with a verb and occur between the argument phrases; (2) WOE parse (Weld & Wu, 2010), which uses bootstrapping from entries in Wikipedia info-boxes to obtain extraction patterns in dependency parses; and (3) Ollie (Schmitz et al., 2012), which identifies open-pattern templates over a set of high-precision seed tuples from REVERB and then applies these pattern templates at extraction time.

Helping the machine understand the language is a complex and difficult comprehensive task. Open information extraction is only part of machine reading. There are other linguistic analysis tools for studying the various emotional, cognitive, and structural components in texts. For example, the Linguistic Inquiry and Word Count (LIWC) (Chung & Pennebaker, 2007) is a validated program for text analysis. LIWC comprises a predefined dictionary clustering the number of instances of nearly 4,000 keywords into 80 psychologically meaningful dimensions. It has been used widely in the social science researches, such as gender differences in language use (Newman et al., 2008), bloggers' personality (Yarkoni, 2010), online deception (Toma & Hancock, 2012) and so on.

For the present study, we adopt OpenIE to extract causal tuples in news reports and answer the following research questions:

RQ1: What are the major causes and consequences of autism mentioned in newspaper coverage?

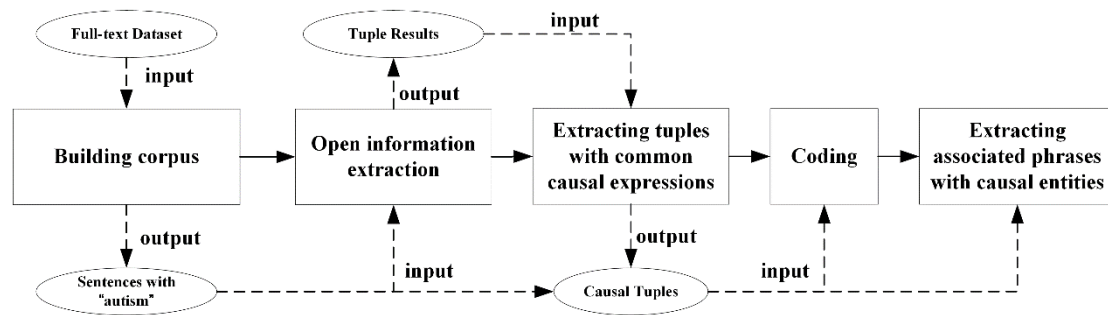
RQ2: How does the media coverage of the causes and consequences of autism vary across time and media sources?

RQ3: What are the main characteristics of reporting styles (e.g., tones) associated with the mentioned causes and consequences of autism?

### **3 Methods**

To reveal the major causes of autism mentioned in news, we adopt a five step

methods to find the casual relations. Figure 1 summarizes the main process of the research methods. First, we download the full-text data of the selected newspapers and extract the sentences containing autism information. Second, we use open information extraction to extract relation tuples from the corpus. Third, we filter out irrelevant tuples with a set of words that express causal relationships. Fourth, we manually select the correct tuples and determine whether the causal relationship is positive or negative by reading its original sentence. Finally, we use co-occurrence frequency and adjacency distance to find the associated phrases with the causal agents.



**Figure 1. Main process of the research method.**

### 3.1 Media Selection

Four major newspapers are used in this study: three American publications, *New York Times*, *Washington Post*, *USA Today*, and one from the UK, *The Guardian*. To select newspapers, we consulted prior research focusing on newspaper coverage of autism and recent classification of the political orientations of western media (Blake, 2014). For instance, Clarke (2008) examined six newspapers (e.g., *New York Times*, *the Guardian*) from the U.S. and U.K., and content analyzed how they covered the vaccine-autism controversy. Among the selected newspapers for the present study, the *New York Times* has been perceived as a relatively liberal newspaper, whereas the *Washington Post* is listed as both conservative (Grefenstette, Qu, Shanahan, & Evans, 2004; Puglisi, 2011) and liberal (Gerber, Karlan, & Bergan, 2009). *USA Today* appeals to a wider range of audiences, and is considered more moderate or centrist (Dennis, 1997), while *The Guardian* is traditionally liberal in perspective, claiming to have a social liberal or social democratic line (The Guardian News and Media Limited, 2015).

#### Dataset

For this study we use the search query “autism or autistic” to identify full-text news articles from the LexisNexis database. This database is regarded as the most authoritative and comprehensive collection of English-language media publications. A few studies (e.g. McKeever, 2012) have used articles from this database to study media coverage of autism.

Most of the downloaded articles we use are already tagged for publication type. To concentrate on news discussing autism, we removed those with any one tag in the following list: Web Log; Text; Summary; Special Report; Series; Schedule; Book

Review; Quote; Question; Paid Death Notice; Series; Op-Ed; Obituary (Obits); List; Letter; Interview; Editorial; Correction; and Biography. The remaining types include: News, News Analysis and Brief. After screening, we got 6504 news articles in total covering 1977-3-31 to 2015-12-31 (see Table 1).

**Table 1. Number of news articles and time coverage for each media**

News Media	# of articles	# after screening	Time of Coverage
USA Today	681	452	1989-1-6 to 2015-12-15
Washington Post	3514	1635	1977-3-31 to 2015-12-24
The Guardian	3262	2513	1984-11-19 to 2015-12-31
New York Times	3644	1904	1980-6-4 to 2015-12-30
<b>Total</b>	11101	6504	1977-3-31 to 2015-12-31

We then used the NLTK’s recommended sentence tokenizer to parse each article into sentences and extract those containing the words “autism” or “autistic.” The procedure resulted in a total of 14,305 sentences that were further subjected to causal knowledge extraction.

### 3.2 Open Information Extraction

We adopt the Open IE 4.1 system to extract the relation tuples. This system is an open source project developed by the University of Washington’s Turing Center. Based on the Open IE 4.1 system, TextRunner Search is a knowledge search engine that extracts over five billion text tuples from over a billion web pages. It can be used to help people understand general questions like “What kills bacteria” or “What contains antioxidants?” (Fader et al., 2011).

Open IE system extracts a great diversity of tuples consisting of argument phrases from the input sentence and a phrase from the sentence that expresses a relation between the arguments, in the format (Entity1; Relation; Entity2). The method is based on detecting and analyzing noun phrases (for entities) and verb-centric phrases (for relations). First, it uses a set of over 110,000 high precision seed tuples to bootstrap a large training set. Second, it learns open pattern templates over this training set. Next, the system applies these pattern templates at extraction time. The extracting process is done without a pre-specified set of relations and with no domain-specific knowledge engineering.

We ran Open IE on the autism sentences dataset, with examples provided of the extraction results (Table 2).

**Table 2. Examples of the Open IE results**

<b>Sentence</b>	As recently as a few decades ago, psychiatrists thought autism was caused by a lack of maternal warmth.
<b>Tuple 1</b>	(psychiatrists; thought; autism was caused by a lack of maternal warmth; T: As recently as a few decades ago)
<b>Tuple 2</b>	Context (psychiatrists thought): (autism; was caused; by a lack of maternal warmth)
<b>Sentence</b>	The notion that autism was caused by “refrigerator” mothers and absent fathers.
<b>Tuple 1</b>	(autism; was caused; by “refrigerator” mothers and absent fathers)



### 3.3 Common Causal Expressions in the News

Girju and Moldovan (2002) validated a list of causation verbs (e.g. “give rise to”) by using a large collection of English news articles from major newspapers such as *Wall Street Journal*. The list of causation verbs is ranked by their frequency of appearance and ambiguity of meaning in the texts. For the present study, we combine Girju and Moldovan’s list with other causation verbs (e.g. “make,” “allow”) identified from a random sample of the studied news articles. This list of more than 100 causation verbs used herein is applied to the studied articles to locate linked entities, with one of them specified as “autism” or its analogous keywords such as “autistic.” To fully capture the causal relations in the news, we use the words of this term list to extract tuples from the Open IE results.

In the Open IE results, the same words in different relational phrases may also have various forms, such as {thimerosal; was causing; autism} and {thimerosal; contributing to; autism}. So in the first step, we use word lemmatization to remove inflectional endings only, returning to the base form of a word. For example, {thimerosal; was caused; autism} and {thimerosal; contribute to; autism} will be transformed into {thimerosal; be cause; autism} and {thimerosal; contribute to; autism}. This makes it leads to more clearly structured easy for us to merge the results after merging, conducive to interpretation. Specifically, the NLTK Wordnet-Lemmatizer is applied to the relation phrases to convert the words into their dictionary form.

After the extraction, we get a total of 2,772 tuples. However, we still face two problems. One is that some of the relations are ambiguous, in the sense that they express a causation relation only in a particular context and only between specific pairs of entities (e.g. {no medical test; develop for; autism}). The other problem is the entities may be expressed in different ways but have the same meaning (e.g. “MMR” and “Measles, Mumps and Rubella”). So in the second step, we manually identify all correct extractions from the tuples and classify different phrases according to medical characteristics. For example, "lack of maternal warmth" and "bad mothering" both belong to parenting, "depression" and "aggression" belong to the symptoms. We also label the “not cause” relation with a tag of “negative” by reading the original sentences (Table 3). The coding procedure is implemented together between the coauthors and an outside expert with autism expertise. We followed an iterative procedure and any discrepancy of selection was discussed and resolved among the collaborators. It should be noted that if a causal tuple has been repeatedly mentioned in an article, we only count it once. Likewise, if an article contains multiple causal tuples, the frequency of each is increased once. Finally, we get 1,512 causal tuples.

**Table 3. Examples of the Coding Results**

<b>ID</b>	<b>Entity 1</b>	<b>Relation</b>	<b>Entity 2</b>	<b>Causal Entity Classification</b>	<b>Relation Type</b>	<b>Date</b>	<b>Media</b>	<b>Original Sentence</b>
1	head banging	be associated with	autism	Symptoms	positive	1994-8-2	Washington Post	<i>But if, for example, your child seems withdrawn and emotionally detached, that may be a sign that the head banging is associated with autism, depression or other underlying.</i>
2	autism	be caused by	“cold refrigerator” mothers and weak, absent fathers	Parenting (Environmental factor)	positive	1999-12-28	New York Times	<i>Fifty years ago, researchers believed that autism was caused by “cold refrigerator” mothers and weak, absent fathers.</i>
3	autism	be caused by	a biological malfunction in the brain	Brain (Biological factors)	positive	2000-3-21	Washington Post	<i>Researchers suspect that autism is caused by a biological malfunction in the brain and are working to try to understand how and where the malfunction occurs.</i>
4	vaccine	cause	autism	Vaccine	negative	2008-7-1	Washington Post	<i>Federal officials and the scientific establishment have steadily maintained that the decision to concede the Poling case does not in any way suggest that vaccines cause autism.</i>
5	MMR vaccine	trigger	autism	Vaccine	positive	2009-10-1	The Guardian	<i>Ashton compared the current furore to the MMR scare, when many parents refused to have their children immunised after Dr. Andrew Wakefield claimed the MMR vaccine might trigger autism.</i>
6	gene	cause	autism	Gene (Biological factors)	positive	2011-8-15	USA Today	<i>He says scientists have identified genes that cause 15% to 16% of autism cases and probably will find more problems.</i>

In order to assess the reliability of the results, we randomly select 100 news articles from the dataset and employ two postgraduates with a background in communication to encode the autism causation tuples in the article. After the coding process, they compare the results with each other. If there is a difference, they need to discuss and resolve discrepancy. Finally, we compare the manual coding results with the tuples extracted with OpenIE. It shows that the recall rate through the machine learning can reach 77% and the accuracy rate is 100%.

There are two reasons for the relatively low recall rate. The first is that not all causal relationships appear in sentences using “autism” or “autistic”, and secondly, because the vocabulary we use does not contain all the causal relationship terms. However, this semi-automated approach has a very high accuracy, and greatly saves research time and improve work efficiency. In the face of the growing network data, the machine reading method has obvious advantages to traditional content analysis, both in terms of exploring complex causality and studies across a long-time span.

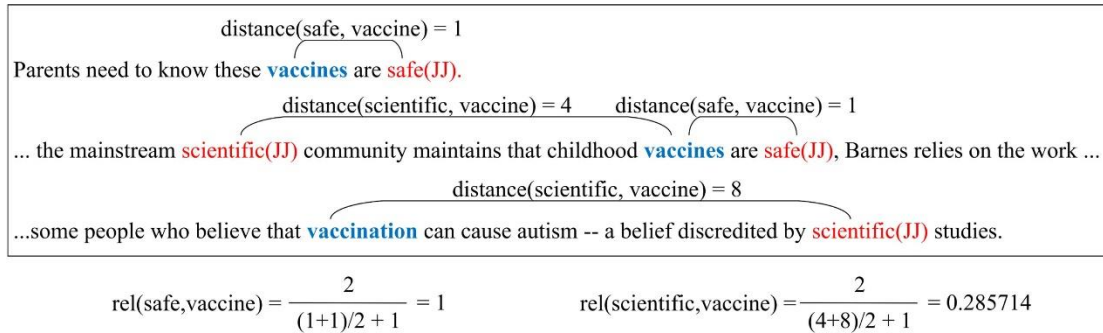
### 3.4 *Extracting associated phrases with the causal agents*

To better understand how media portray different causal entities, we extract the adjective terms in the entities’ related sentences. There are three major causal agents of autism we focus on, which are vaccine, gene, and parenting, as well as the symptoms, regarded as consequences of autism.

First, we use a group of related terms for each agent (Figure 4) to extract the sentences containing these terms from the whole corpus. Then, we use NLTK’s part-of-speech tagger (POS-tagger) to find adjectives in this data. Here, two different taggers are adopted, UnigramTagger and BigramTagger. These two taggers can extract adjectives in one-term forms and phrases with two terms. After that, we use the co-occurrence frequency ( $frq(adj, agent)$ ) and adjacency distance ( $avg\_distance(adj, agent)$ ) to measure the relation ( $rel(adj, agent)$ ) between the adjective and causal agent, which can be expressed as:

$$rel(adj, agent) = \frac{frq(adj, agent)}{avg\_distance(adj, agent) + 1}$$

The co-occurrence frequency refers to the number of times that the adjective and causal agent-related words appear in the same sentence. The adjacency distance is the average number of words between each pair of these two words. The bigger the co-occurrence frequency and smaller adjacency distance, the stronger relationship between the phase pairs. An example of computing the relation score of (safe, vaccine) and (scientific, vaccine) is presented in Figure 2.



**Figure 2. Example of the process for calculating the relation score between the adjective word and the causal agent.**

Finally, we sort the phrases according to the relation scores and keep the top 20. By counting the frequency of them, we then remove the total number of less than five for further analysis.

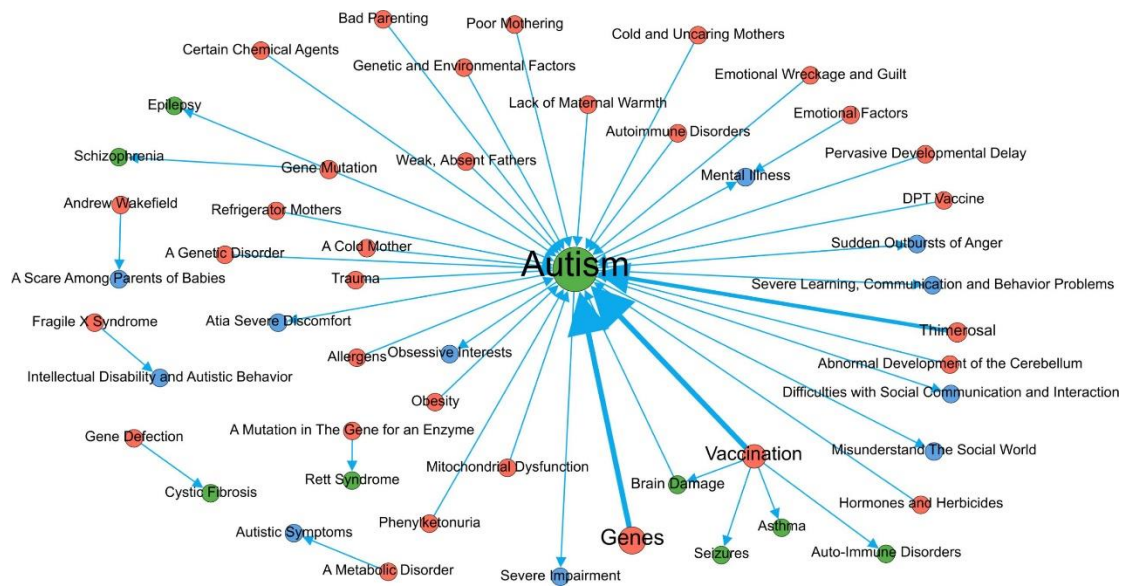
## 4 Results and Discussions

Table 4 shows the results of the total number of news articles reporting the causal relationship with different relation terms. The results show that only about 23.2% of articles in all autism-related reports directly report the etiology of autism, while most do not directly describe the causation. Tania et al. (2004) examined 627 newspaper articles featuring or citing 111 papers published in 24 scientific and medical journals. Only 11% of the newspaper articles were categorized as having moderately to highly exaggerated claims; the majority were categorized as having no claims (63%) or slightly exaggerated claims (26%).

**Table 4. Total number of causal relationship for each relation term**

Relation term	Number of news articles	Relation term	Number of news articles
cause	397	contribute (to)	54
lead (to)	179	trigger	32
link (to)	175	result (from)	27
develop	173	reduce	22
associate (with)	103	stem (from)	13
bring (up/on)	100	spark	8
relate	80	implicate (in)	5
create	74	provoke	4
produce	66		
<b>Total Number of news articles</b>		1512	

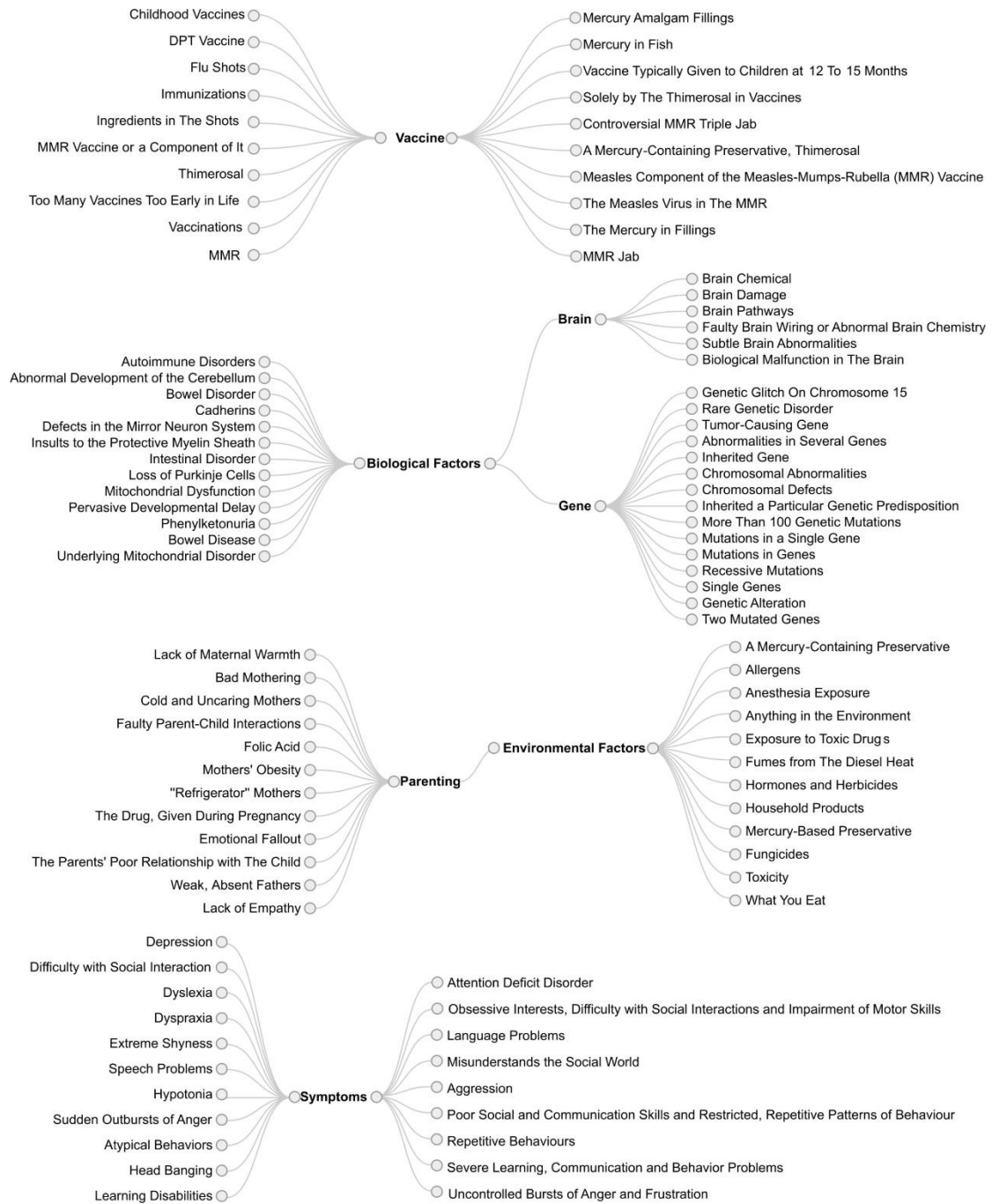
### 4.1 RQ1: What are the major causes and consequences of autism mentioned in newspaper coverage?



**Figure 3. An overview of the major causes and consequences of autism in newspaper coverage.**

Based on the newspaper data used in this study, the results of research question 1 show that genes, vaccination, and poor parenting are considered to be the major causes of autism. In Figure 3, each solid dot represents an entity, with red dots representing the causes of autism and blue dots representing the impacts or symptoms of autism. There are also some diseases (green dots) frequently mentioned with autism in the news. The size of the dot denotes the frequency of that entity appearing in causal statements in media. The size of the direct arrows denotes the frequency of that particular causation being noted in media coverage.

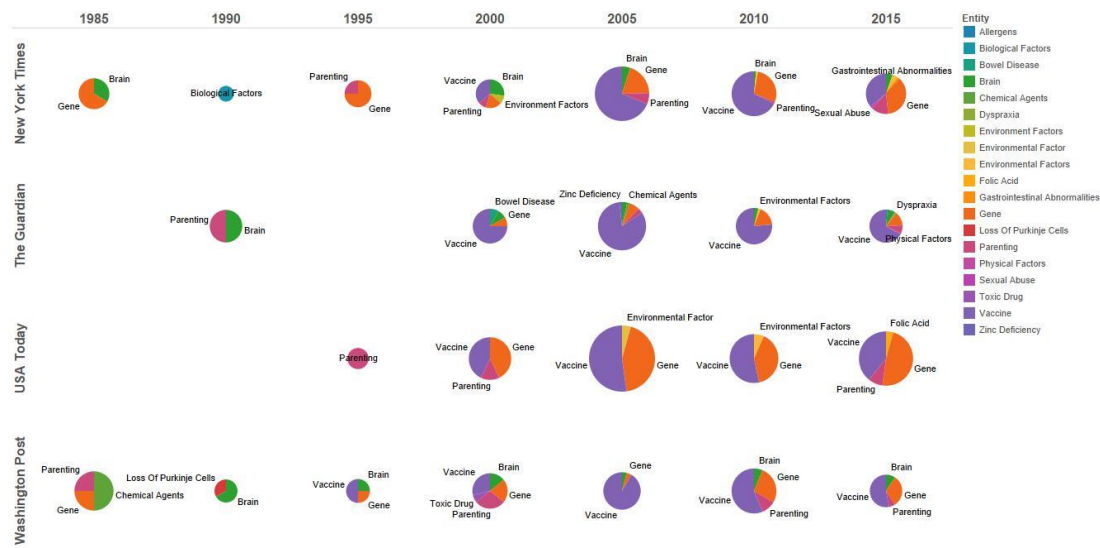
The graphical trees in Figure 4 show the different expressions and connotative breadth of each major entity thought to cause autism. “Vaccine” is more associated with mercury, MMR, thimerosal. “Gene” is tightly associated with genetic defects, mutation, and inheritance. Matson and LoVullo (2009) reviewed more 10,000 articles on autism in scientific journals, and observed that genetics is the most frequently studied topic. “Parenting” is more focused on the mother in the family, a common feature of “autism blame.” Indeed, the term “refrigerator mother” suggests how important a mother’s affective response to the child mitigates autism and its symptoms. In Leo Kanner’s work (1943) that first identified autism, he proposed that autism may be related to a lack of warmth among the fathers and mothers. However, he rejected the “refrigerator mother” theory later because only some of the parents of autistic children have that characteristics. More importantly, the media report this relationship as a causality, which is actually a distortion and error. The reason why we place parenting under environmental factors is mainly because that parenting is an external factor, rather than an internal factor, causing autism.



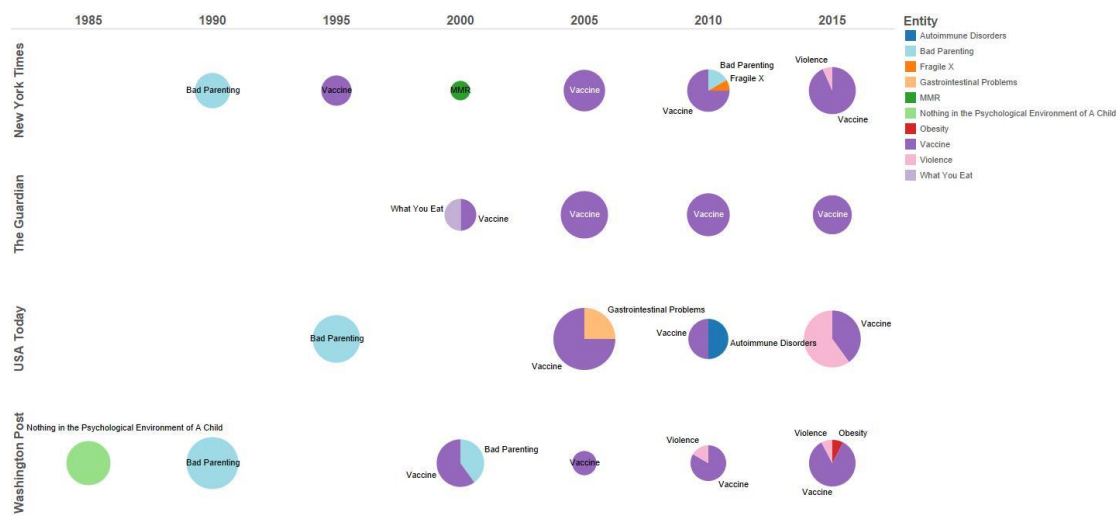
**Figure 4. Different expressions for the major causes of autism mentioned in news**

**4.2 RQ2: How does the media coverage of the causes and consequences vary across time and sources?**

Next, we plot these causal factors according to their coverage time, the number of reports and the overall proportion for each period of a 5 years' window.



**Figure 5. Report rate of media-supported causal factors of autism across different sources (in a five-year window).** The size of the circle represents the number of news reporting the casual factors during each five years. The circle at the beginning denotes the total number before 1985. The colored sectors represent the proportions of different causal factors.



**Figure 6. Report rate of media-denied causal factors of autism across different sources (in a five-year window).**

As shown in Figure 5, across time, the media-supported causal factors of autism have become more diverse. The two factors of Parenting and Gene have been mentioned as causal factors much earlier as compared to other factors such as Vaccine and Environmental Factors. It is clearly shown that Vaccine in relation to mercury exposure via thimerosal is a major focus of media coverage of autism causation during 2001 and 2010. In the past five years, however, the factor of Gene has become the strongest focus of media coverage, with Vaccine as secondary.

Reporting on the causes of autism is closely related to the definition and

conceptualization of autism put forth by various sources. The early psychogenic model of autism, rooted in psychiatric theories, is characterized as a form of parent blaming. Research following this model mostly subscribes to the general argument that autism develops in infancy in response to the emotionally distant, absent, or cold parents. Autism was later conceptualized as a neurodevelopmental disorder, emphasizing biological causes rather than parental behavior. Alternative conceptualizations of autism recast the disorder in a biomedical framework, offering hopes to cure certain forms of autism via medicine. One frequently discussed biomedical factor is special diet (Baker, 2008). Baker labeled the two lines of conceptualizations as “mainline” (parent and neurological) and “alternative” (biomedical) thinking.”

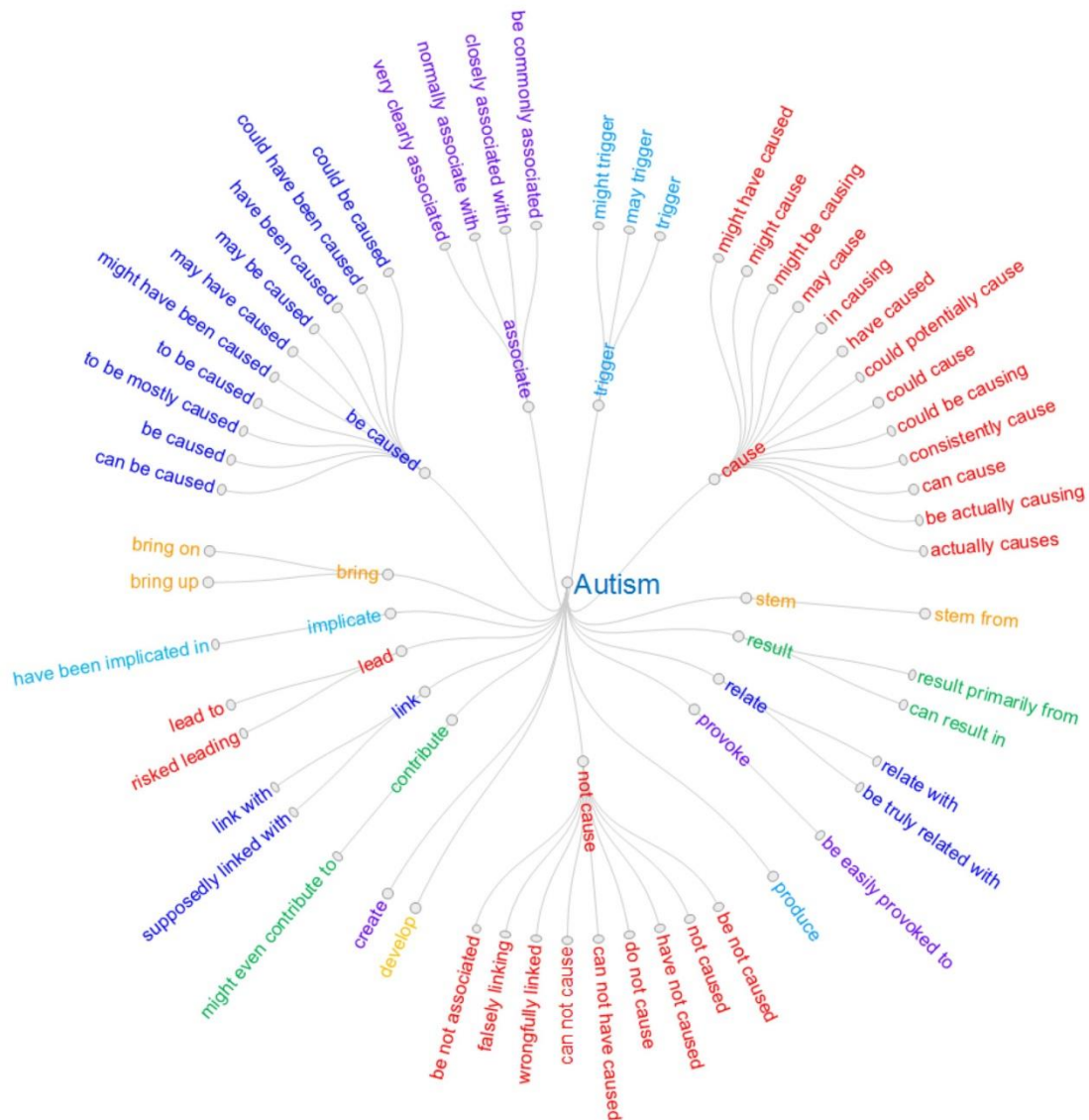
Figure 6 shows which factors have been denied by many parties (but not all) as causes of autism. Clearly, the two main negated factors in the media are Parenting and Vaccine. According to the history of evolving conceptualization of autism, parenting corresponds to the early psychogenetic model stressing the impact of affective coldness (e.g., “refrigerator mother” and absent father) in causing autism. Denial of vaccination as a cause became more salient after 2000. Even so, reports of autism caused by vaccines are still much more than those that deny this causal relationship. This phenomenon also points to the lack of media attention to tracking scientific discovery of a particular disease etiology. Also, it speaks to the unbalanced reporting of different media platforms which can seriously affect the audience’s opinion. Dixon and Clarke (2012) employed experimental methods to examine the effects of media reporting on the autism-vaccine controversy on public perceptions of vaccine safety and behavioral intentions. The authors found that falsely balanced reporting (i.e. presenting information for or against the autism-vaccine link) significantly increased respondents’ uncertainty about vaccine safety, as well as lowered their intention to vaccinate their children in the future. The authors further reported that failing to present scientific consensus on no linkage between autism and vaccine is a major factor biasing public perceptions of the scientific community in favor of or against their opinions. This can lead to very serious social problems, as shown in Figure 2, “Andrew Wakefield” caused “a scare among parents of babies”.

The Washington Post and the New York Times reported the etiology of autism much earlier than the other media. After 2000, the USA Today is significantly higher than other newspapers in terms of the number of reports that support or deny some etiology. Claims of both pro-link and anti-link between vaccine and autism were identified in the newspaper coverage. Such a finding is consistent with prior research, as the journalistic norm of balanced reporting is deep-trenched in western media (Clarke, 2008). More concretely, by presenting conflicting views of the link between vaccine and autism, media have rendered themselves to be neutral and professional on a controversial topic, at least on the facade.

How news stories describe causality can also affect the opinion of audiences. Media framing theory identifies the media's ability to manipulate interpretations of a media message through careful control of angles, facts, opinions, amount of coverage



(Entman, 1993). Figure 7 presents more detailed information about the words that reporters used to describe the causes of autism, where they employed a relatively tentative tone (using words such as “might,” “could,” “may”) to describe these causes. Previous studies have shown that, news articles that featured health-related research and non-health-related research yielded significant differences in terms of tone. Health-related research articles were less critical than articles featuring non-health-related research (Racine et al., 2006). Here, this “safe” expression for media to describe the etiology of autism once again confirmed the discovery. After all, the real cause of autism has not been fully confirmed.

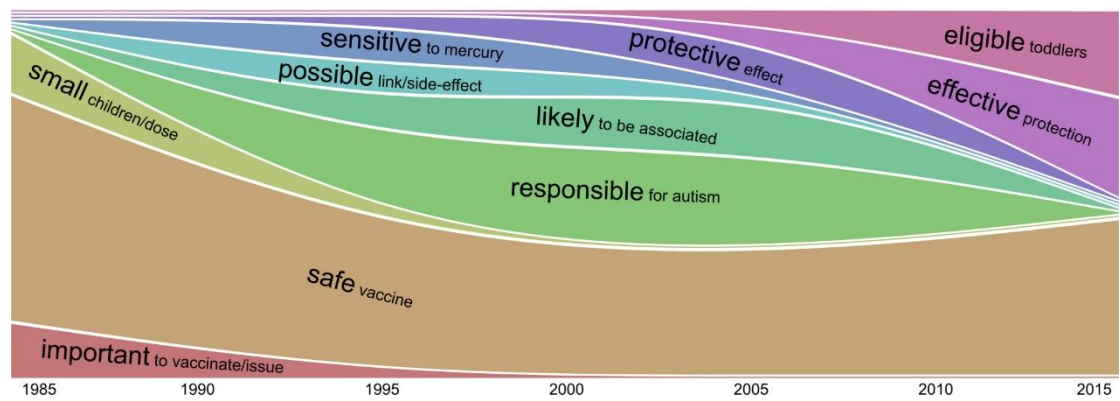


**Figure 7. Words used to describe the causes of autism in major newspapers.**

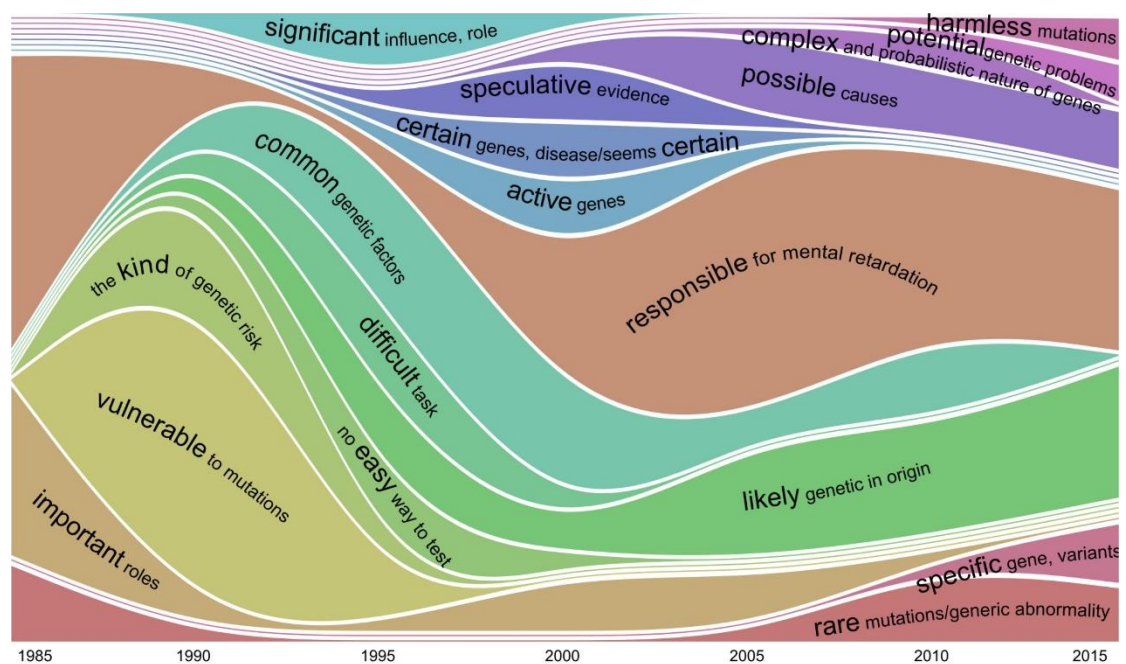
**4.3 RQ3: What are the main characteristics of reporting styles (e.g., tones) associated with the mentioned causes and consequences of autism?**

Research question 3 is addressed in results reflected in Figure 8(a-d). In order to demonstrate how the articles describe the entities, we further expand the extracted

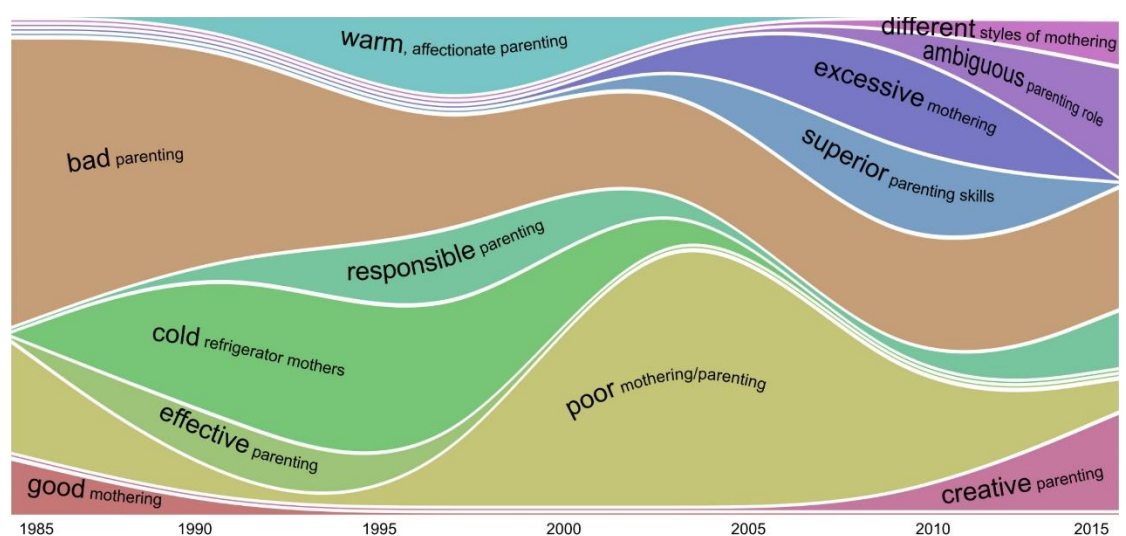
adjectives to the associated phrases.



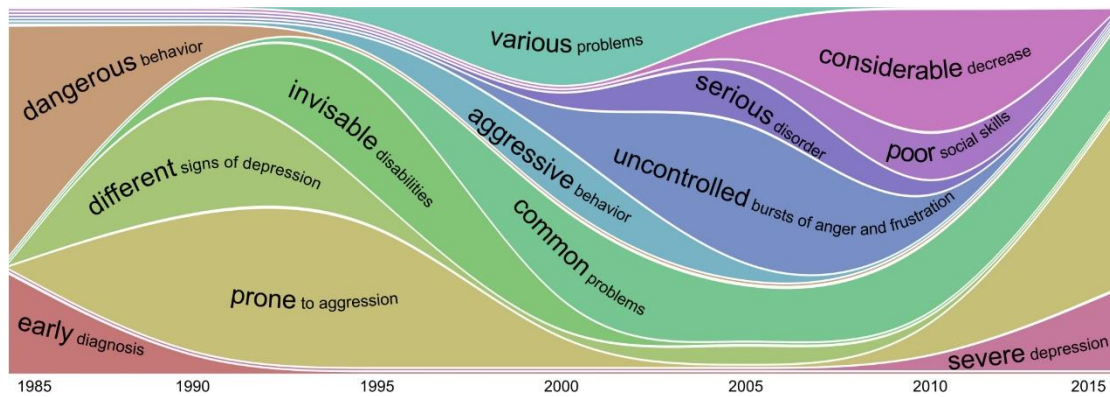
(a) Vaccination



(b) Genetic factors



(c) Parenting



(d) Symptoms

**Figure 8. Thoughts and emotional phrases associated with the media-mentioned causes and consequences of autism.** The different colors in each river plot indicate different adjective phrases, and the variation of the width of each river over time shows changes in the frequency of the phrase.

We can see from Figure 6 (a), in the past 30 years, the media have been trying to correct the public awareness about the vaccine will lead to autism. Voices for supporting and opposing this assertion have been constantly fighting for the right to speak. Eventually, reports stating “safe vaccine” prevailed.

The media have never denied the effects of genetic factors, but emphasized that confirming genes lead to autism is an important task by using the words like "difficult task", "no easy way to test" and "complex". At the same time, the rising frequency of words such as "likely" and "possible causes" in recent years has shown that more news, instead of highlighting the point that the genes will lead to autism, has preserved enough space for the public and scientists to discuss this assumption.

Although the “refrigerator mother” has no longer appeared in the news reports, the discussion of bad parenting may cause autistic children has never stopped. During the last 10 years, the parenting skills have attracted enormous attention from media.

The word "dangerous behavior" disappears after 1995. One possible reason for this is that such a description may lead to a result that the public alienated and stigmatized autistic patients, and even hurt them. More symptoms are now described as saying that autistic patients themselves suffer great pain, such as "severe depression", "poor social skills" and "serious disorder", calling for social support and attention for this vulnerable group.

## 5 Conclusion

Overall, this study reveals a complex landscape of media coverage of autism causation. Concretely, the media have presented multiple potential causes and symptoms of autism over the past decades. Despite scientific advances, media

coverage in general has been tentative or inconclusive about the causes of autism and its symptoms.

It should be pointed out that, in scientific literature, the debate about autism causation has never abated. For example, the controversy of vaccination and its potential linkage to autism continues to this day, as in 2016 California joined Mississippi and West Virginia in mandating vaccines for all children in school. But California's Senate Bill 277, put in place after outbreaks in measles, whooping cough, and other disease-related deaths that are found to be prevented by vaccines, has caused people to move out of state or home school to avoid vaccination (*The Mercury News*, 2016). Baker (2008) reviewed the history of the thimerosal controversy, and found that public mindset on the autism epidemic was formed and shaped by three different but interlinked narratives about vaccine preservatives, mercury poisoning, and its contribution to autism. This is largely due to anecdotal data provided by parents whose children appeared to exhibit extreme reactions to vaccines, which in many cases include conditions such as autism, asthma and seizures.

The present study shows that the major media coverage of autism causation does vary considerably across different media sources. So-called liberal media such as the *New York Times* appears to be quick in keeping pace with the advances in science. For instance, genetic research has been more frequently covered in this newspaper. In contrast, relatively conservative newspapers such as the *Washington Post* have been more frequently covering those "wedge" issues such as vaccination. This media coverage also reflects the current challenges scientific researchers have been encountering in addressing in autism study. For instance, research has not identified specific genes that are definitely related to autism. Such a state of uncertainty is well captured in those words of a speculative nature versus concrete assertions, found across all media reports.

There are still some limitations in this study. First, the dataset only contains sentences with "autism" and "autistic", and the causal conjunctions is fixed, which might narrow the data sample and potentially bring bias to the results which we have discussed in the method evaluation section. In the future, we will try to improve the machine reading technologies to extract the causal relationship. Secondly, the relationship between the adjective and the entity cannot be completely determined by the co-occurrence. Meanwhile, the emotions associated with different causal entities are only analyzed by the description, without further detecting the sentiment orientation of the reports.

Overall, the current study serves an exploratory example for applying machine reading to inspect media coverage of scientific causation issues. The implication of this work can improve the public understanding of how journalism reports modern scientific discoveries, especially those related to health information. More importantly, it facilitates the use of computational linguistics in analyzing the diversity of media roles and reporting forms. With the help of open IE, scientists can effectively reduce the size of the research data and focus on the valid information for

answering questions. Even in hard science, it has never been simple to establish a causal relationship in addressing many complicated diseases, including autism. Media, on this matter, should be cautious with covering such intricate and contested research and relationships. Outside of media reports, future research can also investigate how other types of texts such as scientific publications, government documents, and social media present issues pertaining to autism causation.

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

## References

- Baker, J. P. (2008). Mercury, vaccines, and autism: one controversy, three histories. *American Journal of Public Health*, 98, 244-253.
- Blake, A. (October 21, 2014). Ranking the media from liberal to conservative, based on their audiences. Washington Post. Retrieved from [https://www.washingtonpost.com/news/the-fix/wp/2014/10/21/lets-rank-the-media-from-liberal-to-conservative-based-on-their-audiences/?utm\\_term=.821e954c2f32](https://www.washingtonpost.com/news/the-fix/wp/2014/10/21/lets-rank-the-media-from-liberal-to-conservative-based-on-their-audiences/?utm_term=.821e954c2f32)
- Brown, P., Zavestoski, S. M., McCormick, S., Mandelbaum, J., & Luebke, T. (2001). Print media coverage of environmental causation of breast cancer. *Sociology of Health & Illness*, 23, 747-775.
- Bubela, T. M., & Caulfield, T. A. (2004). Do the print media “hype” genetic research? A comparison of newspaper stories and peer-reviewed research papers. *Canadian Medical Association Journal*, 170(9), 1399-1407.
- Clarke, C. (2008). A question of balance: The autism-vaccine controversy in the British and American elite press. *Science Communication*, 30, 77-107. doi: 10.1177/1075547008320262
- Chung, C., & Pennebaker, J. W. (2007). The psychological functions of function words. *Social communication*, 343-359.
- Cronin, K. A., van Leeuwen, E. J. C., Mulenga, I. C., & Bodamer, M. D. (2011). Behavioral response of a chimpanzee mother toward her dead infant. *American Journal of Primatology*, 73(5), 415-421.
- Currenti, S. A. (2010). Understanding and determining the etiology of autism. *Cellular and molecular neurobiology*, 30, 161-171.
- Dennis, E. E. (1997). How “Liberal” Are the Media, Anyway? The Continuing Conflict of Professionalism and Partisanship. *Harvard International Journal of Press/Politics*, 2(4), 115-119.
- Diagnostic and Statistical Manual of Mental Disorders (DSM-5) (2013). Retrieved from <https://www.psychologytoday.com/blog/my-life-aspergers/201203/is-autism-becoming-more-common>
- Dixon, G., & Clarke, C. (2012). The effect of falsely balanced reporting of the autism-vaccine controversy on vaccine safety. *Health Education Research*, 1-8.
- Downs, M. (2016). WebMD Special Report. Retrieved from <http://www.webmd.com/brain/autism/searching-for-answers/vaccines-autism>

- Entman, R. M. (1993). Framing: Toward clarification of a fractured paradigm. *Journal of communication*, 43(4), 51-58.
- Etzioni, O. (2007, October). Machine reading of web text. In *Proceedings of the 4th international conference on Knowledge capture* (pp. 1-4). ACM.
- Fader, A., Soderland, S., & Etzioni, O. (2011, July). Identifying relations for open information extraction. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing* (pp. 1535-1545). Association for Computational Linguistics.
- Ferrucci, D. A. (2012). Introduction to "this is watson". *IBM Journal of Research and Development*, 56(3.4), 1-1.
- Gerber, A. S., Karlan, D., & Bergan, D. (2009). Does the media matter? A field experiment measuring the effect of newspapers on voting behavior and political opinions. *American Economic Journal: Applied Economics*, 1(2), 35-52.
- Gernsbacher, M. A., Dawson, M., & Goldsmith, H. H. (2005). Three reasons not to believe in an autism epidemic. *Current directions in psychological science*, 14(2), 55-58.
- Ghosh, S., Shankar, N., & Owre, S. (2011). Machine reading using markov logic networks for collective probabilistic inference. In *Proceedings of ECML-CoLISD, 2011*.
- Girju, R., & Moldovan, M. (2002). Text mining for causal relations. *Proceedings of the FLAIRS Conference*, 360-364.
- Grefenstette, G., Qu, Y., Shanahan, J. G., & Evans, D. A. (2004, April). Coupling niche browsers and affect analysis for an opinion mining application. In *Coupling approaches, coupling media and coupling languages for information retrieval* (pp. 186-194). LE CENTRE DE HAUTES ETUDES INTERNATIONALES D'INFORMATIQUE DOCUMENTAIRE.
- Kanner, L. (1943). Autistic disturbances of affective contact. *Nervous Child*, 2, 217-250.
- Landrigan, P. J. (2010). What causes autism? Exploring the environmental contribution. *Current opinion in pediatrics*, 22(2), 219-225.
- Matson, J. L., & LoVullo, S. V. (2009). Trends and topics in autism spectrum disorders research. *Research in Autism Spectrum Disorders*, 3(1), 252-257.
- McKeever, B. W. (2012). News framing of autism: Understanding media advocacy and the combating autism act. *Science Communication*. doi: 10.1177/1075547012450951
- Mercer, L., Creighton, S., Holden, J. J. A., & Lewis, M. E. S. (2006). Parental perspectives on the causes of an autism spectrum disorder in their children. *Journal of Genetic Counseling*, 15, 41-50. doi: 10.1007/s10897-005-9002-7
- Miller, V. A., Schreck, K. A., Mulick, J. A., & Butter, E. (2012). Factors related to parents' choices of treatments for their children with autism spectrum disorders. *Research in Autism Spectrum Disorders*, 6, 87-95.
- Newman, M. L., Groom, C. J., Handelman, L. D., & Pennebaker, J. W. (2008). Gender differences in language use: An analysis of 14,000 text

- samples. *Discourse Processes*, 45(3), 211-236.
- Newschaffer, C. J., Croen, L. A., Daniels, J., Giarelli, E., Grether, J. K., Levy, S. E., ... & Reynolds, A. M. (2007). The epidemiology of autism spectrum disorders\*. *Annu. Rev. Public Health*, 28, 235-258.
- Parrott, R., Silk, K., Weiner, J., Condit, C., Harris, T., & Bernhardt, J. (2004). Deriving lay models of uncertainty about genes' role in illness causation to guide communication about human genetics. *Journal of Communication*, 54(1), 105-122.
- PBS (2002). Refrigerator mothers: History of autism blame. Retrieved from <http://www.pbs.org/pov/refrigeratormothers/fridge/>
- Phillips, D. P., Kanter, E. J., Bednarczyk, B., & Tastad, P. L. (1991). Importance of the lay press in the transmission of medical knowledge to the scientific community. *New England Journal of Medicine*, 325, 1180-1183.
- Poon, H., & Domingos, P. (2010, July). Unsupervised ontology induction from text. *In Proceedings of the 48th annual meeting of the Association for Computational Linguistics* (pp. 296-305). Association for Computational Linguistics.
- Poon, H., & Vanderwende, L. (2010, June). Joint inference for knowledge extraction from biomedical literature. *In Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics* (pp. 813-821). Association for Computational Linguistics.
- Puglisi, R. (2011). Being the New York Times: the political behavior of a newspaper. *BE Journal of Economic Analysis & Policy*, 11(1).
- Racine, E., Bar-Ilan, O., & Illes, J. (2006). Brain imaging A decade of coverage in the print media. *Science communication*, 28(1), 122-143.
- Schmitz, M., Bart, R., Soderland, S., & Etzioni, O. (2012). Open language learning for information extraction. *EMNLP-CoNLL '12 Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, 523-534.
- Schreck, K. A., Russell, M., & Vargas, L. A. (2013). Autism treatments in print: media's coverage of scientifically supported and alternative treatments. *Behavioral Interventions*, 28(4), 299-321.
- Singhal, A. (2012). Official Google Blog: Introducing the knowledge graph: Things, not strings. Retrieved from <https://googleblog.blogspot.co.za/2012/05/introducing-knowledge-graph-things-not.html>
- The Guardian News and Media Limited (2015, July 26). The Scott Trust: values and history. *The Guardian*. Retrieved from <https://www.theguardian.com/the-scott-trust/2015/jul/26/the-scott-trust>
- The Mercury News (2016). California's vaccine law: Opponents moving, home schooling to avoid controversial mandate. Retrieved from <http://www.mercurynews.com/2016/06/30/californias-vaccine-law-opponents-moving-home-schooling-to-avoid-controversial-mandate/>

- Toma, C. L., & Hancock, J. T. (2012). What lies beneath: The linguistic traces of deception in online dating profiles. *Journal of Communication*, 62(1), 78-97.
- Verhoeff, B. (2013a). Autism in flux: A history of the concept from Leo Kanner to DSM-5. *History of Psychiatry*, 24, 442-458.
- Verhoeff, B. (2013b). The autism puzzle: Challenging a mechanistic model on conceptual and historical grounds. *Philosophy, Ethics, and Humanities in Medicine*, 8(1), 1.
- Weigold, M. F. (2001). Communicating science: A review of the literature. *Science Communication*, 23, 164-193.
- Weld, D. S., & Wu, F. (2010). Open Information Extraction using Wikipedia. *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, (July), 118–127.
- Yarkoni, T. (2010). Personality in 100,000 words: A large-scale analysis of personality and word use among bloggers. *Journal of research in personality*, 44(3), 363-373.