Issues in Information Systems
Volume 22, Issue 3, pp. 184-196, 2021

DOI: https://doi.org/10.48009/3_iis_2021_200-213

Text mining to identify the origin of chronic wasting disease

G. Kent Webb, *San Jose State University, g.webb@sjsu.edu*

Abstract

Chronic wasting disease (CWD) is a 100 percent fatal, prion disease of deer that has the potential to decimate the deer population and jump to the human population. A simple Google search on the "origin of chronic wasting disease" yielded a total of 56 relevant articles. Of these, 75 percent report that the origin is unknown, 19.6 percent report that the disease may have originated in a Fort Collins, Colorado, government research facility, and 5.4 percent report other possible origins. Government sources reported the Fort Collins theory 4.3 percent of the time while non-governmental sources, such as news articles, reported the Fort Collins theory 32.1 percent of the time. Text mining of the internet for the first 40 years of the disease produced evidence supporting a common assertion in the press that all of the early cases can be traced back to Fort Collins. For 1967 into 1998, six clusters were identified that could all be traced back to Fort Collins. Limited information from game farms made tracking difficult for 1998 to 2007 with 10 more clusters traced back to areas linked to Fort Collins or with trace backs to Fort Collins explainable based on the evidence. Available at: www.deerfriendly.com/deer-disease/chronic-wasting-disease/possible-origins-of-chronic-wasting-disease

Keywords: Text Mining, Contact Tracing, Analytics/Business Intelligence

Introduction

Early information about the threat of COVID-19 to the United States came from a doctor working for the state of California who saw Twitter reports of Chinese being locked in their homes. A related group of doctors began translating web pages in China reporting death results (Dickerson, 2021). A significant increase in the recent use of social media in medical research was noted by Zhang et al. (2020) who used text mining to identify research themes. In 2020, Zhou reported on a multi-modal text mining project to create a repository of information on COVID-19 to facilitate research. The origin of that disease has been a major topic for discussion.

A disease with a 100 percent fatality rate, but proceeding at a much slower pace than COVID-19 and endemic to the deer population, was first observed in January, 1967, at a Fort Collins, Colorado, research facility operated by Colorado State University and the Colorado Division of Wildlife. It was identified and dubbed chronic wasting disease (CWD) when observed in 1978 at the Wyoming Game and Fish Department's research center north of Laramie which had reportedly traded deer with Fort Collins. It is generally believed to be a prion disease, a misfolded protein, similar to scrapie in sheep, mad cow in cattle, and Creutzfeldt-Jakob disease in humans. As a test of the potential transmission to humans, squirrel monkeys were "inoculated with brain tissue from a CWD-infected mule deer" (Marsh, et al., 2005). The monkeys were euthanized starting at 31 months after the inoculation when they developed a progressive neurodegenerative disease similar to CWD.

This paper applies text mining to evaluate the most commonly cited explanation for the disease origin, that it jumped to deer from sheep with scrapie in shared pens at the Fort Collins facility – the Fort Collins theory. Contact tracing for the first 40 years of the disease using information from the text mining process supports the Fort Collins theory, but does not rule out other possible explanations. In reviewing the results, it appeared that government agencies were less likely than non-governmental sources to report the Fort Collins theory and instead report that the origin is unknown. A formal hypothesis test was applied to a sample from a Google search that verified this tendency.

Text mining and websites for disease information collection and distribution

While data mining typically involves structured data, text mining involves unstructured textual data (Tan, 1999). Hassani (2020) identifies text mining as an emerging, powerful tool. Ye (2016) proposed a text mining methodology using tools such as logistic regression to divide public cancer research documents into different categories for easier reference. Wang and Lo (2021) give a summary of text mining automation tools applied to the fast growing research related to COVID-19. Loque et al. (2020) provide a broad review of tools for text mining for medical applications. Hoa et al. (2018) provide a general survey of the literature on text mining in medical research and observe a substantial expansion of literature on the topic. Text mining of websites has been widely used as noted by Steinberger (2008) who describes its increased use among government agencies in Europe searching the internet for information that might identify a threat to public health.

Strzelecki (2020) measures a decrease in the number of sites with medical information showing up in Google search, proposing that the company has raised standards in ranking these sites given the potential danger of bad medical information. Eschenfelder and Miller (2007) attempt to develop a metric to value websites publishing information about chronic wasting disease, but find many different value measures to consider. Several websites make effective use of graphical computer simulation to model the risk and transmission of CWD. Farnsworth et al. (2007) describe how a Geographical Information System at a fine landscape level can support understanding of transmission risk. Devivo et al. (2017) present a model suggesting CWD is already responsible for declining deer herds in Wyoming. A site that is often at the top of a Google search about CWD, 'cwd-info.org' provides a searchable archive of important events in the disease history and provides summaries and links to recent news stories.

Røyeng (2020) comments on the ephemeral nature of the internet with broken links and changing technologies, and proposes a collaborative internet archive for personal and social use. A recent book by Devendran and Arunkumar (2020) gives great detail on the practical issues related to archiving the Web as attempted at 'web.archive.org' with more than 338 billion web pages archived. An archiving project that ran through 1999 and is still on the Web at www.mad-cow.org contains information used by this project in web pages about chronic wasting disease, a related prion disease. A goal of this project is to archive deer management information.

Case Example: Chronic Wasting Disease (CWD)

The text mining process described here began in 2010 as a general search to find and conserve deer information for research and management decisions. Having cast a wide net over many years for information, and given the fleeting nature of some content on the internet, the process saved important information that was not retrievable at a later time when this more systematic search began.

An important source for the most commonly cited explanation for the origin of the disease, the Fort Collins theory, comes from a newspaper interview with a master's degree candidate conducting a nutrition experiment with deer at Fort Collins during the 1967 period where he reports that deer were penned with sheep from a scrapie project, a related prion disease (Gerhadt, 2001). As a test of the theory of a sheep origin for the disease, researchers using laboratory experiments concluded "it is surprising that CWD cases have not been reported elsewhere in the world where cervids and scrapie-infected sheep coexist" (Tamgüney et al., 2009). Gerhardt's article was preserved on the internet by the site DeerFarmer.com or it would not have been discovered by this text mining project.

The incubation period for CWD can range from about 15 to 34 months. As a result and with limited testing, the disease has often already been in an area for several years or more before it is detected. Testing currently requires killing the deer. Transmission is thought to be by direct contact through saliva, feces, urine, or blood and by indirect contact with contaminated food, water, or soil. Since prions are not living organisms, they are hard to destroy. One method is to apply heat at 900 degrees Fahrenheit for several hours which has led to suggestions that forest fires be started in infected areas. As of January, 2021, the Center for Disease Control reports that CWD is found in wild deer, elk, or moose in at least 25 states and two Canadian provinces. The exchange of animals among captive herds has been the primary wide-scale disease distribution mechanism.

As analyzed in a following section, a simple Google search on the "origin of chronic wasting disease," done in the spring of 2021, resulted in a total of just 56 relevant articles. The search provided a snapshot of what the public would have seen if investigating this issue. Government sources were much less likely to report on the Fort Collins theory than non-governmental sources in this simple Google search, although the first two pages of the Google search results did provide information on the major different theories. The long-term text mining process for this research turned up hundreds of relevant articles. Several news articles from reliable organizations asserted that all of the early infections could be traced back to Fort Collins, but the evidence for all the trace backs did not appear in any source among the thousands of sources that were reviewed. The primary goal of this research was to conduct and document the contact trace back to Fort Collins.

For the 16 clusters of chronic wasting disease (CWD) identified in the United States over the first 40 years of the disease, 1967 to 2007, approximately 400 sources were used to create a summary graphic of the contact tracing that appears in a following section. As illustrated in the graphic, the text mining discovered evidence to trace back the first six clusters, 1967 into 1998, to Fort Collins. Of the 10 remaining clusters detected from 1998 to 2007, five clusters can be traced back to an infected area linked to Fort Collins and five clusters have evidence that supports and explains a trace back to an infected area linked to Fort Collins. Poor record keeping in some captive game facilities has been a notable problem throughout the disease history. The evidence for the contract tracing is summarized in a following table. Example sources are listed. Nearly all trace backs have multiple sources of support. All of the evidence is available at www.deerfriendly.com/deer-disease/chronic-wasting-disease/possible-origins-of-chronic-wasting-disease. Several sources speculated that the disease could be traced back even further to when sheep with scrapie from Europe were imported into the United States.

Methodology

The text mining process used for this research is illustrated in Figure 1. Starting in 2010, Google Alerts, a service that automatically searches the internet for keywords on a time frequency specified by the user, were set up to provide daily reports for about 56 categories that were identified as common search topics

by using Google trends, another service giving a graphical summary of Google keywords search volume. The keyword search volume indicated that searches were commonly done by state, accounting for 50 categories, and for general topics such as population, suburban management, and disease. Later, a daily Bing search was added after it was discovered that Google Alerts was missing some important information (Webb, 2016). As important topics were discovered, new categories were created on separate web pages and targeted search was used to find supporting information. In this case, the Google search option to set a custom date range was helpful. For example, search may have turned up an important event from a previous time and this option allowed a targeted search for that time period. The text mining process used for this research is illustrated in Figure 1

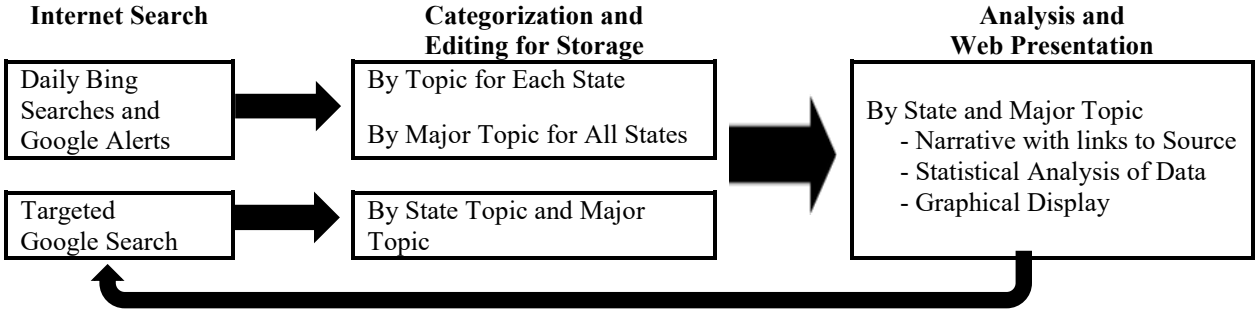

Figure 1: The Text Mining Process

Most of the results from the search were discarded because they were not on topic, redundant, or not containing information thought to be significant. Useful, non-copyrighted documents and snippets of information from copyright sources were stored on a public website by state and by major topic. Approximately 21,000 documents or snippets were collected for the general topic of "deer" and about 4,000 for the topic of "chronic wasting disease." In reviewing the collected information, it seemed that government sources were less likely to report on the Fort Collins theory than non-governmental sources such as news organizations or websites created by non-governmental sources. Below is a statement of this observation as a hypothesis test.

H_{a1}: Government sources are less likely to report the Fort Collins origination theory than non-governmental sources.

As a test, a simple Google search on the "origin of chronic wasting disease" was done in spring, 2021, with the results presented in Table 1. The research hypothesis is accepted with a p-value below 0.05. Whether a Google search can be considered to be a random sample is explored by Bar-Yossef and Gurevich (2018) who test for a bias toward long articles, but find none. Although a statistical test is applied in Table 1, the results can also be considered to be a population of what a user would have found using Google search to explore this issue. The major hypothesis of this research, that the early cases of CWD can be traced back to Fort Collins, is also supported, but in a qualitative analysis.

H_{a2}: The early cases of chronic wasting diseases can be traced back to the government research facility in Fort Collins, Colorado.

Results

Of the 170 results from the simple Google search, 56 directly addressed the question of first origin for the disease. As presented in Table 1, there is a significant difference in the likelihood that scrapie from sheep

at Fort Collins as an origin for chronic wasting disease is reported by a private news or other non-governmental (NGO) organizations compared to a government source. Governmental sources included state wildlife agencies, the National Institutes of Health, and the United States Department of Agriculture. The category of Other for the reported origin includes spontaneous generation, exposure to toxins, or scrapie from another location. Table 1, created in SPSS, uses subscripts on the cell numbers to identify statistically different row values at the 0.05 significance level.

Chi-square = 6.736. Two-sided asymptotic significance of 0.149 With Scholarship removed as a source, chi-square = 6.410, significance is 0.041. N = 56		Source of Information						Total	
		News and NGO		Government		Scholarship			
		N	%	N	%	N	%	N	%
Reported Origin	Unknown	18_a	64.3%	20_a	87.0%	4_a	80.0%	42	75.0%
	Fort Collins	$\mathbf{9_a}$	32.1%	$\mathbf{1_b}$	4.3%	$1_{a,\,b}$	20.0%	11	19.6%
	Other	1_a	3.6%	2_a	8.7%	0_a	0.0%	3	5.4%
Total		28	100%	23	100%	5	100%	56	100%

Table 1: H_{a1}, Hypothesis Accepted. Cross tabulation of the Reported Origin of chronic wasting disease against the Source of the Information: News and other Non-governmental Organizations (NGO), Government, and Scholarship (articles published in scholarly journals). The two bold values in the table indicate a statistically significant difference at the .05 level with News and NGO more likely to report that deer exposed to scrapie at Fort Collins may have been the origin of chronic wasting disease than government sources. Assumes an independent sampling process.

Table 1 represents the population of what a user would have found in doing a search for the origin of chronic wasting disease in spring, 2021. As the hypothesis test confirms, the government sources were less likely to report the Fort Collins theory. However, the Fort Collins theory appeared on the first page of the search results. Another theory, that deer already exposed to sheep and infected with scrapie were brought into Fort Collins from the wild, appears on the second page. The Google search presented a reasonable spectrum of the prevailing theories on the first two pages.

Among the scholarly papers identified in the search, one mentioned Fort Collins as a good explanation for the origin in a discussion reviewing the literature and history of the disease. Four of the five scholarly articles simply stated the origin was unknown or uncertain in a brief statement without delving much into the background of the disease. One scholarly article was coded as a government document because it was published by research scientists working for the government laboratory at Fort Collins or for Wyoming Department of Game and Fish. They suggest spontaneous creation north of Fort Collins as the likely source. Scrapie is often mentioned in the article, but not as a source related to Fort Collins (Miller et al., 2000). The paper rests its conclusion on the observation that the geographic center of the disease was north of Fort Collins.

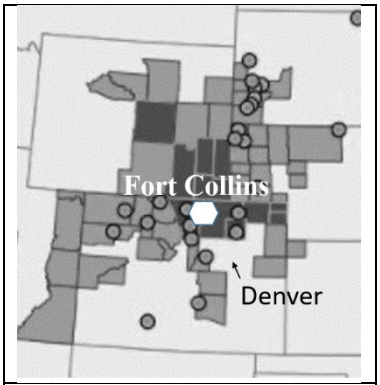

Figure 2: From 2007 USGS Infection Map for CWD

Figure 2 shows the location of Fort Collins, the white hexagon, against a map of chronic wasting disease (CWD) infections as of 2007 prepared by the United States Geological Survey (USGS) (Richards, 2007). The center of the infection map, calculated north to south through Fort Collins, is roughly 50 miles to the north of Fort Collins. An inspection of a satellite map for the area shows highways and significant urban development to the south and southeast of Fort Collins which is located just to the north of the Denver Metropolitan area. Those structures would impede deer movement to the south

as modeled by Robinson et al. (2013). They note that spread is strongly influenced by highways and rivers that block deer movement. Figure 2 is a small section around Fort Collins taken from the larger map in Figure 3 which explains that the dark areas on the map indicate where CWD was discovered before 2000. The dots represent infected captive herds. When Miller et al. (2000) did their study, testing had only discovered the disease in the darkened area which would have placed the known area of the disease farther to the north.

The Fort Collins facility was releasing some deer back into the wild. Assuming that those deer would be released outside the suburbs and into deer habitat, they would necessarily be released north of Fort Collins. According to reports, Fort Collins was also trading deer with the Laramie, Wyoming, research facility that was north of Fort Collins. Given these circumstances, it should be expected that the disease center would be north of Fort Collins. Also, it is much more likely the disease jumped species where deer were sharing pens with sheep, as at Fort Collins, than out on the open range where deer may have encountered sheep.

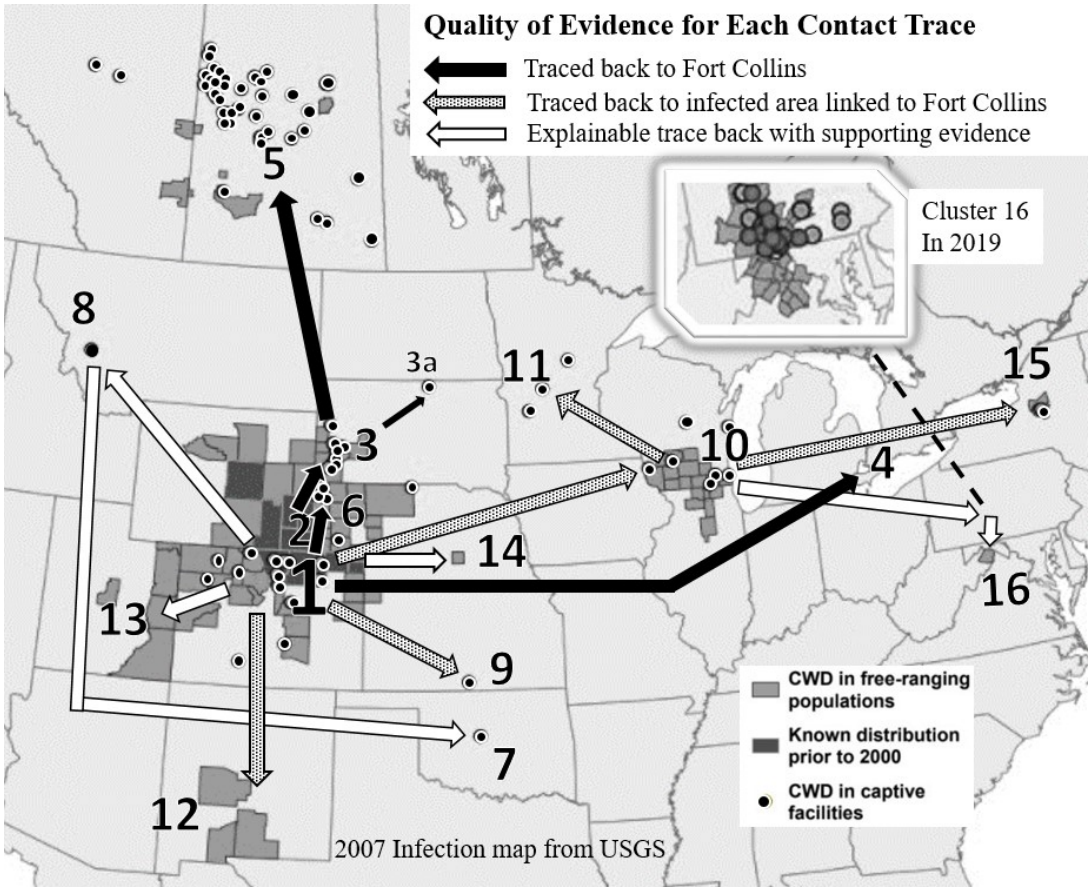

Figure 3: Contact tracing the first 40 years of chronic wasting disease, 1967 to 2007. USGS (Richards, 2007) map of infections overlaid with infection clusters numbered in chronological order and disease contact tracing from text mining. Cluster 4, the Toronto Zoo, was not included in the USGS map.

Among the many news articles saved as part of the text mining, a few asserted that all early cases of CWD could be traced back to Fort Collins. No record showing all of the details was discovered. To address this issue, the map in Figure 3 was developed to show the trace back routes for 16 clusters of the disease that were identified from the first observation in early 1967 to early 2007. In 2007, the United States Geographical Survey (USGS) released a CWD fact sheet with a map showing infections by county and by

captive herd (Richards, 2007). Figure 3 shows the mapped infections overlaid with arrows showing the results of the contact tracing, starting with Fort Collins, and the level of evidence supporting each identified route. Table 2 provides details.

Table 2: Chronic Wasting Disease Contact Tracing, 16 Clusters, 1967 to 2007

Cluster	Location	Detected In Wild	Detected Captive	Details
1	Colorado	1981	Observed 1967	Observed at a Fort Collins, CO, government research station in 1967 where reportedly deer and elk shared pens with sheep from a scrapie project. Some deer and elk were released into the wild; others were shipped to facilities such as the Denver Zoo. First identified in the wild in a Colorado elk in 1981.
2	Wyoming	1985	Identified 1978	Identified as chronic wasting disease in 1978 in a Wyoming State Veterinary Lab that had traded animals with Fort Collins.
3 3a	South Dakota	2001	1996-97	Detected in 7 captive elk farms with elk from the Colorado and Wyoming infection zone during the winter of 1996-97, including the McPherson County farm (3a). First detected in wild, white-tail deer in 2001 in Fall River County near infected elk farms.
4	Toronto		1978	An isolated case in the Toronto Zoo. Deer had been transferred from the Denver Zoo which had received deer from Fort Collins. No escapes.
5	Saskatchewan and Alberta	2000	1996	Detected in 1996 in Saskatchewan among farmed elk imported from an infected farm in South Dakota. Detected nearby in wild deer in 2000.
6	Western Nebraska	2000	1998	Infected elk in a game farm confirmed in 1998 originated from an infected Colorado herd, some through a Montana game farm. First discovered in wild deer in 2000 within 3 miles of an infected farm.
7	Oklahoma		1998	Infected elk originated from infected herd in Montana
8	Montana		1999	Detected near Phillipsburg in 1999 in an elk farm that was active in the elk trade. Poor records, lots of unexplained elk deaths, suspected of taking wild elk from infected areas. Consensus that it came from another elk farm.
9	Kansas		2001	Found in a captive bull elk in Harper County in 2001 that came from a Colorado elk farm.
10	Wisconsin, Illinois	2001	2002	Three wild deer taken in the 2001 hunt. The first deer farm to test positive was in 2002 in Walworth County. A deer that had escaped from an infected farm was found in March, 2002. Deer farms had been importing captive deer from Colorado, Wyoming, and Saskatchewan since 1996.

Table 2 is continued on the next page

Clusters in the map are numbered in chronological order with number one being Fort Collins. Cluster details were taken from public sources explaining how the disease had spread in that area. Much of the spread is associated with the small black dots that indicate infected captive herds. At Fort Collins, deer were released into the wild at the end of projects. Deer often escape from captivity so that the first infections in wild deer are often found near captive herds. Deer are routinely exchanged among captive herds, transporting the disease. Good examples of this process are cluster 3 and 3a in South Dakota with elk from the infected area of Colorado and Wyoming distributed among other captive herds including cluster 3a, some distance away. Cluster 5 in Alberta and Saskatchewan, Canada, also illustrate the distribution effect of trading captive animals.

Table 2 (continued): Chronic Wasting Disease Contact Tracing, 16 Clusters, 1967 to 2007

Cluster	Location	Detected in Wild	Detected Captive	Details
11	Minnesota		2002	The first positive test in 2002 in an elk farm in Aitkin County, although the state had done little testing. Other elk farms later tested positive. Wisconsin and Minnesota captive herds had been trading animals.
12	New Mexico	2002		An infected deer found in 2002 at the White Sands Missile base in southern New Mexico where, according to a wildlife manager in nearby west Texas, deer were imported from infected areas of Colorado.
13	Utah	2002		A mature buck from the northeast part of the state near Vernal taken in the fall of 2002 hunt tested positive. Mule deer migrate long distances in this area. Later maps show connection to disease areas in Colorado.
14	Central Nebraska	2004		Discovered in 2004 when scientist began testing along the Platte River, thinking it would be a CWD corridor from western Nebraska. Later testing was consistent with this theory.
15	New York	2005	2005	Discovered in two deer farms and two nearby wild deer in Oneida County in 2005, apparently brought to the area by a taxidermist who disposed of parts of an infected trophy deer from out of the area. Quick containment halted the spread.
16	West Virginia	2005		Discovered in a 2005 road killed deer in Hampshire County just a few months after the first case in New York. Discovered in a nearby Adams County, Pennsylvania, deer farm in 2012. The insert from 2019 in the Figure 3 map shows this area became a large cluster centered around southern Pennsylvania deer farms just over the West Virginia border.
Outside of United States				
	South Korea	2001		Elk imported from Saskatchewan, an infected area, cluster 5.

The text mining processed discovered one infection cluster not included on the USGS map, cluster 4, the Toronto zoo. This finding led to the discovery of an article in the Canadian Veterinary Journal (Dube, et al., 2006) that reported the Fort Collins facility had been taking 'orphan' fawns from the wild to stock their facility. Adult wild deer cannot generally be confined in small pens. These deer were later shipped to other

locations including the Denver Zoo which shipped several animals to the Toronto Zoo. The map with the supporting details in Table 2 is posted at www.deerfriendly.com/deer-disease/chronic-wasting-disease/possible-origins-of-chronic-wasting-disease where links to all the sources are available. Table 3 contains a sample of the sources used to create this history.

Table 3: A sample of key sources for the contact tracing numbered by cluster.

1	Chronic Wasting Disease: Coming to a Deer Population Near You. March 11, 2018. *Sierra Magazine*
2	Don't Let CWD Stop Management. December, 2002. James C. Kroll, *North American Whitetail*
3	Cousin to mad-cow disease hits deer, elk. March 16, 1998. *High Country News*
4	Retrospective investigation of chronic wasting disease of cervids at the Toronto Zoo, 1973–2003. December, 2006. *The Canadian Veterinarian Journal*
5	Chronic wasting disease in Canada: Part 1. May, 2004. *The Canadian Veterinarian Journal*.
6	Montana's game farm industry:| An indictment for abolishment. 2001. Gary R. Holmquist, *The University of Montana*
6	Chronic Wasting Disease (CWD): America's answer to mad cow disease. *mad-cow.org*
7	Montana burns game farm elk. January 31, 2000. *High Country News*
8	Disease is wasting the West's wild herds. September 27, 1999. *High Country News*
8	Chronic Wasting Disease (CWD): America's answer to mad cow disease. *mad-cow.org*
9	No new cases of chronic wasting disease found in Kansas. March 27, 2008. *KDPT*
10	The Killer Among Us. January 31, 2000. *MilwaukeeMag.com*
10	Chronic Wasting Disease in Free-Ranging Wisconsin White-Tailed Deer. May, 2003. *Emerging Infections Diseases*. 9(5). 599-601
11	Diseased Elk Found in Manitowoc County. March 25, 2003. *cwd-info.org*
11	Growing threat: Chronic wasting disease in deer. September 5, 2018. *Winona Post*
12	Chronic wasting disease in deer can be managed. August 20, 2015. *The Times Herald*
14	Deer Harvested Near Grand Island Tests Positive for Chronic Wasting. December 10, 2004. *CWD-INFO.org*
15	Chronic Wasting Disease. November 28, 2016. *Louisiana Sportsman*
15	Remain vigilant of diseases like CWD. November 21, 2018. New York. *Poughkeepsie Journal*
16	Adams County deer farm escapee tests negative for chronic wasting disease, December 7, 2012. *PennLive.com*

For cluster 8 in Montana, the white arrow indicates that even though the game farmer was active in the elk trade across many states, poor records limited trace efforts by the state. The farmer shipped infected elk to other game farms and some of his elk had been quarantined before for tuberculosis. For cluster 10 with a gray arrow into Wisconsin and Illinois, although the first detection was in wild deer, shortly afterwards game farms that had been importing deer from Colorado, Nebraska, and Saskatchewan began to test positive nearby. In 2002, a deer that had escaped from one of the infected farms was found in the wild.

A white arrow, indicating a feasible route with no direct evidence, points to cluster 13 in Utah, an area where mule deer are known to migrate long distances and where some areas are adjacent to infected wild deer in Wyoming. Testing had been limited and often the disease is prevalent before it is discovered. For cluster 14, scientists in Nebraska began testing along the Platte River thinking it would be a conduit for the disease. The white arrow pointing to cluster 14 is based on this assumption, later supported when testing filled in the area between cluster 14 and cluster 6.

At about the same time that an infected deer was apparently brought into New York from some distance away, an isolated wild deer tested positive in northeastern West Virginia. The two white arrows to this cluster 16 indicate no direct evidence for a contact trace, but the Figure 3 map has an insert for this area showing that by 2019 it had become a major cluster centered around nearby southern Pennsylvania game

farms. According to a Pennsylvania Department of Agriculture (2012) news release, when a captive deer tested positive the state quarantined 24 farms in 12 counties that had been trading deer. At the same time there were news reports that deer had escaped from the farms and from an unlicensed farm. As in many states, the disease may have been present there for several years before being detected

Discussion

A simple Google search as discussed around Table 1 provided a fairly representative cross section of theories on the origin of chronic wasting disease. The reported origin, though, varied significantly by information source with non-governmental, such as news websites, more likely to report Fort Collins as a possible origin. Only one of the government sources found in the simple Google search mentioned the Fort Collins theory. As with COVID-19, government agencies can be reluctant to acknowledge potential culpability for releasing a devastating disease.

The text mining process turned up one disease cluster at the Toronto Zoo that had not been included in the USGS map. There was probably also an outbreak in the Denver Zoo. Approximately 400 sources were discovered and stored on the web site that provids the details for the contact tracing. The results support the assertion of some newspaper articles that all of the early cases of the disease can be traced back to Fort Collins. For the 16 clusters in the first 40 years, the text mining process generated evidence supporting the trace back to Fort Collins for the first six clusters, five more clusters could be traced back to infected area linked to Fort Collins, and in 5 clusters the evidence supported an explanation for tracing the disease back to an area linked to Fort Collins. The evidence does not definitively exclude other theories for the disease origin. At minimum, Fort Collins was a primary catalyst in the widespread distribution of the disease.

The text mining approach applied here was labor intensive, but could possibly be duplicated by an artificial intelligence (AI) process looking for trending topics and contradictions in information on the internet. That approach could provide an unbiased fact checker that might significantly reduce the required labor for the many applications of fact checking. The internet provides a big data set for training such an AI. Finding web services for long-term storage of mined information, replacing the longevity of a book, is another challenge for projects with an archival goal.

Ignoring the likely origin of this disease discounts the lax management of captive animals that has been the driving force for this biological disaster. As this paper goes to publication, the state of Texas has imposed an emergency quarantine on 270 captive herds in 95 counties linked to six infected facilities. The state has the nation's largest deer herd.

References

Bar-Yossef, Z. & Gurevich, M. (2008). Random sampling from a search engine's index. *Journal of the ACM*. 55(5), Article number 24, 1-74. https://doi.org/10.1145/1411509.1411514

Devendran, A. & Arunkumar, K. (2020). A Framework for Web Archiving and Guaranteed Retrieval. In: Sharma N., Chakrabarti A., Balas V. (eds). *Data Management, Analytics and Innovation. Advances in Intelligent Systems and Computing*. Vol 1016. Springer, Singapore. https://doi.org/10.1007/978-981-13-9364-8_16

DeVivo, M.T., Edmunds, D.R., Kauffman, M.J., Schumaker, B.A., Binfet, J., Kreeger, T.J., Richards, B.J., Schätzl, H.M., & Cornish, T.E. (2017). Endemic chronic wasting disease causes mule deer population decline in Wyoming. *PLoS ONE.* 12(10). https://doi.org/10.1371/journal.pone.0186512

Dickerson, J. (2021, May 2). Doctors, scientists who warned officials about oncoming pandemic focus of new Michael Lewis book. *CBS, 60 Minutes* broadcast. www.cbsnews.com/news/michael-lewis-premonition-60-minutes-2021-05-02/

Dube, C., Mehren, K.G., Barker, I.K., Peart, B.L., & Balachandran, A. (2006). Retrospective investigation of chronic wasting disease of cervids at the Toronto Zoo, 1953-2003. *The Canadian Veterinary Journal.* 47(12), 1185-1193.

Eschenfelder, K.R., & Miller, C.A. (2007). Examining the role of web site information in facilitating different citizen–government relationships: A case study of state chronic wasting disease web sites. *Government Information Quarterly*, 24(1), 64-88. https://doi.org/10.1016/j.giq.2006.05.002.

Farnsworth, M. L., Hoeting, J.A., Hobbs, N.T., Conner, M.M., Burnham, K.P., Wolfe, L.L., Williams, E.S., Theobald, D.M., & Miller, M.W. (2007). The role of geographic information systems in wildlife epidemiology: models of chronic wasting disease in Colorado mule deer. *Veterinaria Italiana* 43(3), 581-593.

Gerhardt, G. (2005, November 5). Possible origins of chronic wasting disease. *Rocky Mountain News*. Reprinted by permission in DeerFarmer.com. www.deerfarmer.com/wiki/chronic-wasting-disease-origins

Hao, T., Chen, X., Li, G., & Yan, J. (2018). A bibliometric analysis of text mining in medical research. *Soft Computing*. 22, 7875–7892

Hassani, H., Beneki, C., Unger, S., Mazinani, M. T., & Yeganegi, M. R. (2020). Text mining in big data analytics. *Big Data and Cognitive Computing*, 4(1). https://doi:10.3390/bdcc4010001

Luque, C., Luna, J.M., Luque, M., & Ventura, S. (2019). An advanced review on text mining in medicine. *Wires Data Mining and Knowledge Discovery*. https://onlinelibrary.wiley.com/doi/abs/10.1002/widm.1302

Marsh R.F., Kincaid A.E., Bessen R.A., & Bartz J.C. (2005, November). Interspecies transmission of chronic wasting disease prions to squirrel monkeys (Saimiri sciureus). *Journal of Virology*. 79(21):13794-6. https://doi: 10.1128/JVI.79.21.13794-13796.2005.

Miller, W.M., Williams, E.S., McCarty, C.W., Spraker, T.R., Kreeger, T.J., Larsen, C.T., & Thorne, E.T. (2000). Epizootiology of chronic wasting disease in free-ranging cervids in Colorado and Wyoming. *Journal of Wildlife Diseases* 36(4), 676-690.

Pennsylvania Department of Agriculture (2012). Chronic wasting disease not found in Adams county escaped deer. https://www.prnewswire.com/news-releases/chronic-wasting-disease-not-found-in-adams-county-escaped-deer-182416511.html

Richards, B. (2007). Chronic wasting disease fact sheet. *United States Geological Survey*. https://pubs.usgs.gov/fs/2007/3070/pdf/fs2007-3070.pdf

Robinson, S.J., Samuel, M.D., Rolley, R.E., & Shelton, P. (2013). Using landscape epidemiological models to understand the distribution of chronic wasting disease in the Midwestern USA. *Landscape Ecology*. 28(10), 1923-1935. https://pubs.er.usgs.gov/publication/70187413

Røyeng, T. (2020). A collaborative internet archive for personal and social use. *Thesis submitted for the degree of Master in Programming and System Architecture.* Department of Informatics, University of Oslo.

Salloum, S.A., Al-Emran, M., Monem, A.A., & Shaalan, K. (2017). A survey of text mining in social media: Facebook and Twitter perspectives. *Advances in Science, Technology and Engineering Systems.* 2(1), 127-133.

Strzelecki, A. (2020). Google medical update: Why is the search engine decreasing visibility of health and medical information websites? *International Journal of Environmental Research and Public Health*. 17(4):1160. https://doi.org/10.3390/ijerph17041160

Tamgüney, G., Miller, M.W., Giles, K., Lemus A., Glidden, D.V., DeArmond, S.J., & Prusin, S.B. (2009). Transmission of scrapie and sheep-passaged bovine spongiform encephalopathy prions to transgenic mice expressing elk prion protein. *The Journal of General Virology.* 90(Pt 4), 1035-1047. https://doi: 10.1099/vir.0.007500-0.

Tan, A.H. (1999). Text mining: The state of the art and the challenges. In *Proceedings of the Pacific Asia Conference on Knowledge Discovery and Data Mining PAKDD'99 workshop on Knowledge Discovery from Advanced Databases*. 65-70.

Wang, L.L. & Lo, K. (2021). Text mining approaches for dealing with the rapidly expanding literature on COVID-19. *Briefings in Bioinformatics*. 22(2), 781–799. https://doi.org/10.1093/bib/bbaa296

Webb, G.K. (2016). Internet search engine capture success rates and mortality statistics for hyperlinks to public news articles. *Issues in Information Systems*. 17(2), 25-33.

Ye, Z., Tafti, A.P., He, K.Y., Wang K., & He M.M. (2016). SparkText: Biomedical text mining on big data. *PLoS ONE* 11(9): e0162721. https://doi.org/10.1371/journal.pone.0162721

Zhang, Y., Cao, B., Wang, Y., Peng, T., & Wang, X. (2020). When public health research meets social media: knowledge mapping from 2000 to 2018. *Journal of Medical Internet Research*. 22(8). doi: 10.2196/17582

Zhou, X., Mulay, A., Ferrara, E., & Zafarani, R. (2020, October). ReCOVery: A multimodal repository for COVID-19 news credibility research. *CIKM '20: Proceedings of the 29th ACM International Conference on Information & Knowledge Management.* 3205–3212. https://doi.org/10.1145/3340531.3412880