Digital Scholarship in the Humanities, 2023, **00**, 1–13
https://doi.org/10.1093/llc/fqad075
Full Paper

OXFORD

The Victorian anti-vaccination discourse corpus (VicVaDis): construction and exploration

Claire Hardaker[1], Alice Deignan[3], Elena Semino[1,*], Tara Coltman-Patel[1], William Dance[1], Zsófia Demjén[2], Chris Sanderson[1], Derek Gatherer[1,4]

[1]Department of Linguistics and English Language, Lancaster University, Lancaster LA1 4YL, United Kingdom
[2]Institute of Education, University College London, London WC1H 0AL, United Kingdom
[3]School of Education, University of Leeds, Leeds LS2 9JT, United Kingdom
[4]Biomedical and Life Sciences, Lancaster University, Lancaster LA1 4YW, United Kingdom

*Corresponding author. Department of Linguistics and English Language, Lancaster University, Lancaster LA1 4YL, United Kingdom.
E-mail: e.semino@lancaster.ac.uk

Abstract

This article introduces and explores the 3.5-million-word Victorian Anti-Vaccination Discourse Corpus (VicVaDis). The corpus is intended to provide a (freely accessible) historical resource for the investigation of the earliest public concerns and arguments against vaccination in England, which revolved around compulsory vaccination against smallpox in the second half of the 19th century. It consists of 133 anti-vaccination pamphlets and publications gathered from 1854 to 1906, a span of 53 years that loosely coincides with the Victorian era (1837–1901). This timeframe was chosen to capture the period between the 1853 Vaccination Act, which made smallpox vaccination for babies compulsory, and the 1907 Act that effectively ended the mandatory nature of vaccination. After an overview of the historical background, this article describes the rationale, design and construction of the corpus, and then demonstrates how it can be exploited to investigate the main arguments against compulsory vaccination by means of widely accessible corpus linguistic tools. Where appropriate, parallels are drawn between Victorian and 21st-century vaccine-hesitant attitudes and arguments. Overall, this article demonstrates the potential of corpus analysis to add to our understanding of historical concerns about vaccination.

Keywords: vaccine hesitancy; corpus construction; historical data; concordances; collocation

1. Introduction

Vaccine hesitancy is a highly topical phenomenon. In 2019, the World Health Organization described it as 'the reluctance or refusal to vaccinate despite the availability of vaccines' (WHO 2019), and included it among the top ten threats to global health, alongside antimicrobial resistance and climate change. Vaccine-preventable infectious childhood diseases, such as measles, still result in thousands of deaths worldwide every year (Roberts 2020), and the coronavirus disease 2019 (COVID-19) pandemic has seen increased public discussion of vaccination and has highlighted vaccine hesitancy as a public health concern (Puri et al. 2020). Together with this, the internet provides a resource for the dissemination of anti-vaccination arguments (Tafuri et al. 2014), and in recent years, social media discourse has played an increasingly significant role in the anti-vaccination movement (Fajri Nuwarda et al. 2022).

Vaccine hesitancy is not, however, a modern phenomenon (Durbach 2005; Callender 2016), and has existed in an organized form since at least the introduction of the smallpox vaccine in the 19th century (Tafuri et al. 2014). From that point to the present day, vaccine hesitancy in varying guises has been promoted to the wider public, and accompanying misinformation and disinformation campaigns have tended to use the same arguments regardless of the disease and vaccine (Trogen and Pirofski 2021: 498). As was the case for COVID-19, historically, vaccine hesitancy was increasingly articulated alongside government pressure for the public to be vaccinated (Durbach 2005).

There is a good deal of research into historical vaccine hesitancy, much of which uses document analysis. However, until now, there has been no dataset, resource or large body of linguistic data—that is, a corpus—that would permit the kinds of analyses that are possible with, for instance, present-day data and corpus linguistic tools (Coltman-Patel et al. 2022). Corpus linguistics is a method that allows the researcher to analyse extensive textual data in search of

both patterns and fluctuations (McEnery and Hardie 2011). Comparisons can readily be drawn across types of text, over time, between authors and so on. This makes it an ideal approach for the study of evolving vaccine attitudes. As noted above, however, whilst there are collections of historical texts containing discourses around vaccine hesitancy, these did not previously exist in a principled collection suitable for corpus linguistic analysis.

To fill this gap, we developed a 3.5-million-word Victorian Anti-Vaccine Discourse Corpus, hereon 'VicVaDis'. VicVaDis is novel in that it is compiled of ephemeral literature likely to be accessed by the general public, as opposed to texts written for a more educated audience and which have been more fully explored (e.g. Arnold and Arnold 2022). These ephemeral texts are of particular interest to us because, unlike relatively inaccessible scientific literature or complex policy documents or prohibitively expensive formal monographs, they are most likely to have influenced people's decisions on whether to comply with the law and have their children vaccinated. We argue that this characteristic of our corpus makes it comparable to the present-day anti-vaccination discourses that tend to be accessed by the non-specialist public, including via the internet. This article then presents a case study in how corpus techniques can be used to identify the anti-vaccination arguments presented in the corpus.

2. Historical context

As with all data-driven research, it is critical to understand the context in which the data were produced, but this is arguably even more pressing with historical documents that are, themselves, ephemeral in nature and reflect a fast-moving sociocultural climate of uncertainty and scientific advancement. This said, it is impossible for a short historical context section to do full justice to the complexity of 50 years of rapidly evolving legislation, medical science, and social attitudes, let alone to the wider national and international values and discourses that characterized the half century in question. What follows, then, is a very brief summary of the key events and ideas that are most pivotal to our corpus. For more detail on the movements within that period, see Durbach (2005) and for more on legislation, see Williamson (2007).

2.1 Variolation and vaccination

Smallpox was a highly infectious disease which killed at least 30 per cent of those who contracted it (Stewart and Devlin 2006). Following widespread vaccination over a period of more than a century, the WHO declared it eradicated in 1980. The earliest procedure was variolation—the practice of deliberately introducing infectious material from a smallpox pustule to a patient to stimulate an immune response that then gave protection when the disease was later encountered. The procedure has been traced to ancient Asia, and was introduced to the UK in 1721 (Stewart and Devlin 2006). Variolation carried a number of risks, such as developing full-blown smallpox, and the procedure itself had a 2–3 per cent mortality rate (Bonanni and Santos 2011). In 1798, Edward Jenner published *An Inquiry into the Cause and Effects of the Variolae Vaccinae*, which offered a safer alternative to variolation, known as *vaccination*, from the Latin 'vacca', meaning 'cow'. Jenner's innovation was based on a simple instance of folk wisdom: contracting the milder cowpox virus often protected people from contracting the far more contagious and deadly smallpox virus. Vaccination used material known as 'lymph' from the pustules formed on cowpox sufferers, rather than the dangerous smallpox lymph used in variolation. Jenner is thereby credited with saving more lives than any other individual in history (Stewart and Devlin 2006) and his discovery launched a new era in population-level preventative medical practices.

2.2 Vaccination legislation

The period covered by VicVaDis saw a number of landmark legislative acts in England in the area of vaccination, which we now summarize briefly. Before this point, in the early 1800s, persuasion was the only strategy adopted by government and the medical establishment to encourage people to receive vaccinations. In spite of these efforts, the 1837–1840 smallpox epidemic killed 41,644 people in England and Wales. As a result, a series of Vaccination Acts were brought into force, which, for brevity, we number. These began in 1840 with *Vaccination Act 1*. This made the older practice of variolation illegal and punishable, as well as providing free vaccinations for the poor. The Act was relatively ineffectual, however, with less than a third of infants receiving the vaccination (Williams 1994). Under the 1853 *Vaccination Act 2*, if infants were not vaccinated by the age of 3 months, then their parents could be fined or imprisoned. This almost doubled the number of infants receiving vaccination (Williams 1994), but also triggered widespread anti-vaccination sentiment (Tafuri et al. 2014). A test case in 1864 established that hesitant parents could only be fined once per unvaccinated child (Williamson 2007). This, along with a steady decline in infant vaccinations throughout the 1850s and 1860s, led to the passage of *Vaccination Act 3* in 1867, which explicitly allowed enforcement of compulsory vaccinations and repeated penalties.

The 1870–1873 smallpox pandemic resulted in a total of 42,084 deaths in England and Wales, and led to

Vaccination Act 4, which was passed in 1871. The Act obliged each local authority to organize a force of vaccination officers (Durbach 2005). *Vaccination Act 5*, passed in 1874, granted greater control to central government and made it harder for local pockets of vaccine hesitancy to exist within the system. Towards the end of the century, however, the push for mandatory vaccination began to ebb, and in 1889, the Royal Commission[1] on Vaccination was set up. During the 7 years that the Commission sat, many communities suspended prosecutions for non-compliance, leading to further declines in vaccinations. The Royal Commission's final report of 1896 maintained that compulsory vaccinations should continue, but recommended a 'conscience clause' to allow individuals the right to opt out (Swales 1992). This was made law in *Vaccination Act 6*, in 1898. However, invoking this clause required certification by a magistrate, and many applicants had their petitions refused. These difficulties were dealt with by *Vaccination Act 7*, which was passed in 1907. While smallpox vaccination remained technically compulsory in England and Wales for another four decades, *Vaccination Act 7* meant that parents could opt out of vaccination on behalf of their children relatively easily, and many did. This effectively ended the mandatory nature of smallpox vaccination.

2.3 Anti-vaccination sentiment

Resistance to vaccination has existed as long as vaccination itself, and legislation sometimes intensified that feeling (Fajri Nuwarda et al. 2022). *Vaccination Act 3* (1867) in particular galvanized the opposition (Weber 2010), as it resulted in many families in poverty facing repeated, punitive fines. The Anti-Compulsory Vaccination League (ACVL) had formed in 1866 in Finsbury, and, as the further Vaccination Acts were brought into force, its membership began to rise. The original ACVL wound up in 1873 but it was replaced the following year by the National Anti-Compulsory Vaccination League (NACVL). Another organization, the Anti-Compulsory Vaccination and Mutual Protection Society (ACVMPS), offered to pay the fines received by its members for non-compliance, and in 1880 the London Society for the Abolition of Compulsory Vaccination (LSACV) was established.

Resistance to vaccination was not merely formalized through the creation of numerous local and national organizations, however. Growing grassroots resistance to vaccination was increasingly being expressed through civil disorder. In 1853, *Vaccination Act 2* sparked small riots in Ipswich, Henley, Mitford, and various other towns. In 1876, a riot took place in Keighley following the imprisonment of 'the Seven Men of Keighley', Poor Law Guardians who had refused their duty to implement compulsory vaccination to the fullest extent of the law as mandated by *Vaccination Acts 2, 3*, and *4* (Durbach 2005). In 1880, around 10,000 people joined a protest in Dewsbury against compulsory vaccinations, and in 1885, a demonstration took place in Leicester involving around 100,000 people (Charlton 1983; Durbach 2005). As hostility and non-compliance rose, vaccination rates fell to the extent that by 1885, less than half of infants were being vaccinated within the legally mandated three months of birth.

As the unrest became increasingly widespread, towards the end of the century, a new generation of Liberal politicians began to see complete abolition as a vote-winning policy, and in 1896, a NAVL Liberal candidate won a by-election in Reading by the largest ever majority in the borough. By the early 1900s, with the passage of *Vaccination Act 7*, the anti-vaccination campaigners had gained their main objective, in that it had become straightforward for parents to legally avoid having their children vaccinated.

2.4 The anti-vaccination movement's concerns

Existing studies of opposition to compulsory vaccination cite three central grounds: dangers to health, the threat to individual liberty, and religious opposition to the use of animal products. Texts arguing for health dangers cited detailed and gruesome examples of injuries and disease allegedly caused by vaccination, and included photographs of dead children (Durbach 2005: 48), though it is important to note that the after-death photograph was an established genre in the Victorian era, and so this may have been less shocking than it now seems. At a more philosophical level, Durbach (2005) also claims that popular Victorian understandings of health centred around the body being untainted and the blood unpolluted. In this schema, vaccination is not healthy, but rather the reverse, where disease is deliberately introduced into the body. Durbach argues that the anti-vaccination movement had become aligned to similarly motivated and related causes, such as vegetarianism, spiritualism, alternative medicine, the teetotal movement, suffragism, and other alternative causes.

The second reason generally cited for vaccine hesitancy, the threat to individual liberty, is related to wider resentment from working class people at how they were treated, and a mistrust that the governing classes had their best interests at heart (Durbach 2000, 2005). There is some evidence that lawmakers and senior doctors made assumptions about the reasons for hesitancy, attributing it to parental apathy (Durbach 2005; Klondrlik 2020), just one of many negative stereotypes of the poor and working classes.

Finally, religious opposition came from the clergy who saw the introduction of animal matter into humans as a violation of God's will (Fajri Nuwarda et al. 2022).

Moreover, uniting both the health and religious concerns, some people feared that vaccines developed from animal matter might cause them to develop animal features and characteristics (Fajri Nuwarda et al. 2022).

These findings are largely based on manual document analysis of various sections of the surviving literature from the time. We sought to create a corpus that would make it possible to carry out large-scale analyses of the anti-vaccination arguments from that period using corpus linguistic tools.

2.5 Victorian vaccination literature

The Victorian vaccination literature that survives to this day consists primarily of:

- Pamphlets and popular journals produced by anti-vaccination campaigners;
- Local newspaper reports and letters to local newspapers;
- A collection of papers presented to Parliament by John Simon (1857) (e.g. cited by Durbach 2005; Williamson 2007);
- Letters from prominent doctors to various public bodies and publications;
- Medical journals: *The Lancet* and the *BMJ* (analysed by Klondrlik 2020);
- A summary of positions on vaccination in the *Edinburgh Review* (1810), a leading political and literary magazine (analysed by Williamson 2007);
- Materials gathered in the *Monthly Review*, which predated the *Edinburgh Review*, aimed at non-specialist but educated readers (analysed by Arnold and Arnold 2022);
- Contemporary medical histories and textbooks.

Unsurprisingly, more substantial texts with a higher perceived value or official status are more likely to be preserved, and as a result, this body of literature is dominated by specialist tracts and reports aimed at educated readers. Much of the modern research using this collection of texts has sought to study the anti-vaccination movements as political and social phenomena with religious strands, and to situate those movements within their wider historical and social contexts (see, e.g. Williamson 2007; Klondrlik 2020; Arnold and Arnold 2022). However, the skew naturally introduced by the surviving literature has an inevitable impact on our ability to understand the daily lived reality of the ordinary individual from the era and the types of texts and discourses they would have most likely encountered.

In contrast to the educated and establishment texts that were largely produced for specialist and usually privileged audiences, our interests lay more specifically in the anti-vaccination arguments presented to the general public, and we therefore restricted our corpus collection to a subsection of the material that is available. We both excluded content written by specialist and government supported groups, and aimed at a highly educated readership, and strove to include the genres most likely to be read by the general public. Literacy rates increased dramatically in the 19th century in England and Wales, from around 60 per cent of males and 40 per cent of females in 1800, to nearly 100 per cent of both sexes by 1900 (Lloyd 2007). However, most printed material remained prohibitively expensive for working class people (Altick 1957). The most accessible materials were likely to be non-specialist letters in general publications, such as newspapers, popular serialized journals, and pamphlets. Newspaper readership was rapidly growing, to the extent that pubs 'saw newspapers as an attraction worth advertising in their windows, even setting aside valuable space for their reading' (Hobbs 2018: 74). Serialized journals, such as the *Vaccination Inquirer*, were produced by anti-vaccination societies (Durbach 2005) and widely distributed. Pamphlets were cheap or free, and information about print runs suggests that they were often distributed widely (Humphries 2011). We therefore felt reasonably confident that these genres were the most likely to be read by the general non-specialist public.

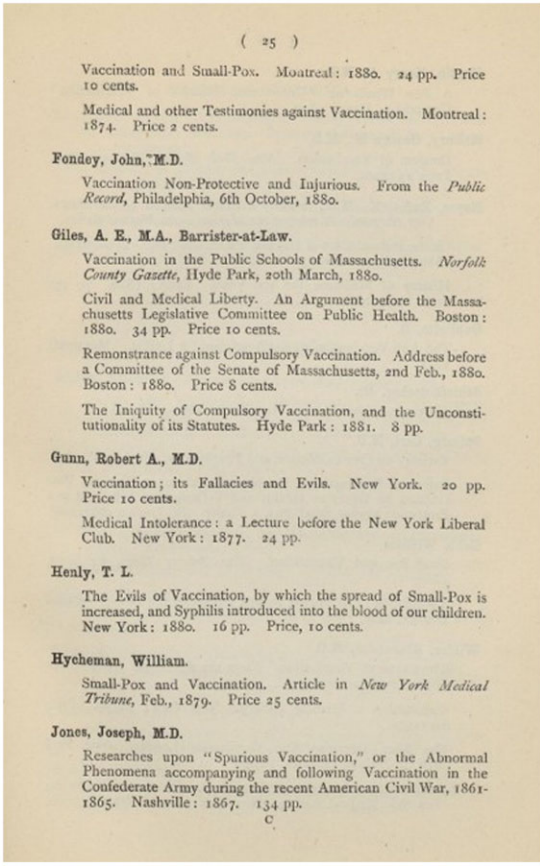
Pamphlets are of particular interest due to their commonalities with modern day social media. They were produced quickly and cheaply, often handed from person to person, and could be written by anyone with the necessary literacy skills (Humphries 2011). The author did not have to be an expert in their chosen topic and might remain entirely anonymous. In the same way, social media posts can be produced quickly and with little or no cost, they are typically designed for public consumption, links can be forwarded from person to person, and the content can be produced anonymously by anyone with the necessary skills, without any requirement for expertise in the chosen topic. During the COVID-19 pandemic, social media has been a major means for the spreading of anti-vaccination sentiments (Puri et al. 2020). VicVaDis thus represents an opportunity for direct investigation of historical anti-vaccination arguments as accessed by the general public, and has the potential to be studied comparatively alongside modern-day social media corpora.

3. The VicVaDis corpus

3.1 Inception and collection

VicVaDis consists of 133 texts dating from 1854 to 1906, and 3,488,959 tokens calculated using the corpus analysis toolkit AntConc (https://www.laurenceanthony.net/software/antconc/). The original inception of

the corpus was the outcome of open-source research into historical publications expressing anti-vaccination sentiments. Having established that there was a reasonably large body of such material, we undertook to develop a principled corpus of anti-vaccination texts. We began by using the search term *vaccination* to query the following archives: (1) the Wellcome Collection Library; (2) Project Gutenberg; and (3) Online Library of Liberty.[2] Among our initial retrieved documents, we found *A Catalogue of Anti-Vaccination Literature* issued by the *London Society for the Abolition of Compulsory Vaccination* (the LSACV) and used this to search for any documents that had not been found in the initial three archives. The LSACV catalogue contains 205 publications by 100 authors, which allowed us to make targeted searches in the UK Medical Heritage Library (UKMHL), the British Library 19th Century Collection, JSTOR, and the Internet Archive.[3] An exemplar page from the LSACV catalogue is shown in Fig. 1.


(25)

Vaccination and Small-Pox. Montreal: 1880. 24 pp. Price 10 cents.

Medical and other Testimonies against Vaccination. Montreal: 1874. Price 2 cents.

Fondey, John, M.D.

Vaccination Non-Protective and Injurious. From the *Public Record*, Philadelphia, 6th October, 1880.

Giles, A. E., M.A., Barrister-at-Law.

Vaccination in the Public Schools of Massachusetts. *Norfolk County Gazette*, Hyde Park, 20th March, 1880.

Civil and Medical Liberty. An Argument before the Massachusetts Legislative Committee on Public Health. Boston: 1880. 34 pp. Price 10 cents.

Remonstrance against Compulsory Vaccination. Address before a Committee of the Senate of Massachusetts, 2nd Feb., 1880. Boston: 1880. Price 8 cents.

The Iniquity of Compulsory Vaccination, and the Unconstitutionality of its Statutes. Hyde Park: 1881. 8 pp.

Gunn, Robert A., M.D.

Vaccination; its Fallacies and Evils. New York. 20 pp. Price 10 cents.

Medical Intolerance: a Lecture before the New York Liberal Club. New York: 1877. 24 pp.

Henly, T. L.

The Evils of Vaccination, by which the spread of Small-Pox is increased, and Syphilis introduced into the blood of our children. New York: 1880. 16 pp. Price, 10 cents.

Hycheman, William.

Small-Pox and Vaccination. Article in *New York Medical Tribune*, Feb., 1879. Price 25 cents.

Jones, Joseph, M.D.

Researches upon "Spurious Vaccination," or the Abnormal Phenomena accompanying and following Vaccination in the Confederate Army during the recent American Civil War, 1861-1865. Nashville: 1867. 134 pp.

C

Figure 1. Example page from a catalogue of anti-vaccination literature.

3.2 Inclusion criteria

The inclusion criteria for VicVaDis were time, location, genre, and technical quality, as follows. We excluded texts published before 1853 and after 1907. Although *Vaccination Act 1* was passed in 1840, it was not until *Vaccination Act 2* in 1853 imposed fines for non-compliant parents that the debate around mandatory vaccines began to rapidly intensify. *Vaccination Act 7* in 1907 signaled the end of meaningful compulsion and provided a natural boundary for the end of our corpus. Given that the 1840 and 1953 acts applied only to England and Wales, documents from Wales could have been included. However, none were found in the archives and catalogue that we queried, therefore the corpus only contains documents published in England.[4]

We only included documents that were anti-vaccination, and excluded works of poetry and fiction. To focus on the anti-vaccination discourse most likely to have been accessed by the general public, we also excluded technical works explaining the legal procedures of the Acts and scientific, academic articles.

Following the rationale described in the previous section, this left us with a collection of pamphlets, newsletters, non-academic tracts and periodicals, and letters to newspapers. We checked the dataset to remove any duplicates, and then excluded any documents where the process of converting the file into a machine-readable text produced a result that fell below a predetermined quality threshold. The majority of our texts had initially been retrieved in PDF format, and needed to be converted to plain text. After trialing several OCR options, we chose Adobe Acrobat. We then created a Perl script to parse the OCR outputs for each document and identify any words that contained internal non-alphabetic characters, such as numbers or punctuation. The real OCR error rate will be higher than the one reported by the script, but the proportion of words with internal non-alphabetic characters allowed us to generally gauge the overall conversion quality of a given text. Using this script, documents with scores of less than 70 per cent were removed. Where examples are provided from the corpus, OCR errors are faithfully reproduced to give a fair sense of the data as it stands, and manual corrections are included in square brackets, e.g. 'its courage and licavenbom [*heaven-born*] principles and convictions'.

We carried out some cleaning of the remaining texts. A second Perl script was written to identify all OCR-introduced errors that occurred more than ten times in the corpus. We then performed a concordance analysis on the results, and this identified 844 distinct true errors which we were then able to correct. We also manually inspected the corpus for visible, repeated errors that the scripts would fail to identify, and

amended them. A full list of all 133 documents in the final VicVaDis corpus is given in Supplementary Appendix A, along with metadata such as each text's filename, year of publication, its full title, declared author,[5] word count, and publisher.

3.3 Corpus composition

As in any corpus, there is a tension between the human-imposed notion of *balance* across, for instance, authors, texts, dates, and so forth, and the more organic principle of *representativeness* (Atkins, Clear and Ostler 1992; Biber 1993). Even in contemporary corpora, objective representativeness is generally an ideal, and in historical corpora such a goal is less achievable still, given the fundamental lack of ground truth. In our case, we have no way of identifying all of the anti-vaccination literature in circulation throughout the 53 years that our texts cover, and some—or possibly, most—of it will be lost without trace. We therefore have no rigorous benchmarks against which to measure our corpus and so cannot know whether particular years, authors, or texts are over- or underrepresented. This has key implications in how the summary statistics around our corpus are understood. We can demonstrate the dispersion of documents over time, between authors, and across texts, but drawing inferences from such results is problematic. We cannot assume for instance, a causative link between the passage of a Vaccination Act in one year and a subsequent spike in the number of texts in the following years. We might speculate that the increase in publications was a direct result of the Act, but we cannot discount the possibility that this result is simply an artefact of a few more texts from that year being preserved, digitized, and made publicly available. It is entirely possible, and even probable, that other relevant texts exist, but that they are simply not retrievable because they have not been digitized, or indexed such that they can be retrieved through standard searches, or released to the public. Through future digitization projects, improved archives, and public releases of private archives, it may therefore be the case that a substantially augmented and updated VicVaDis Corpus becomes viable in years to come. The texts we cannot (yet) access however should not detract from the substantial dataset that we were able to collect, and a summary of this follows next.

3.4 Document characteristics

The shortest text in VicVaDis is a 195-token-letter on the 26 October 1895 to the *East London Observer* entitled *Vaccination and Skin Disease*, in which the writer describes her infant's illness subsequent to receiving the vaccination. The longest text is *The Anti-Vaccinator and Public Health Journal 1872-3*, edited by John Pickering, a 436-page anthology of letters, notes, public addresses, and reports from around the country consisting of 362,864 tokens.

VicVaDis contains sixty-six unique declared authorship designations, ranging from single named authors such as William Tebb (fifty-one unique designations) to anonyms and pseudonyms, such as 'A Sufferer' (six unique designations), to editorial roles and organizations, such as LSACV (five unique designations) to multiauthored publications (four unique designations). The most prolific authors in the corpus are William Tebb and his son, William Scott Tebb. Between them they are responsible for seventeen solo-authored publications ranging from a 489-word letter to the editor of *The Hospital* to a 128,721-word monograph entitled, *A Century of Vaccination and What it Teaches*. After the Tebbs, the next three most prolific declared authors are the LSACV (eleven texts), George S. Gibbs (six texts), and Walter Robert Hadwen (five texts).

Overall, the corpus represents a wide array of authors, but these figures should still be treated with caution. On the one hand, there is a clear disjuncture between the declared and executive authorship on some texts. For instance, we do not know which individual(s) wrote the documents attributed to the LSACV—a problem that affects at least twenty-two of our texts. On the other hand, several texts are edited compendiums and extensive anthologies containing the works of many authors. The result here is that the authorship declared on the texts effectively masks innumerable executive authors contained within the texts themselves.

The chronological dispersion of documents in VicVaDis is shown in Fig. 2.

Over its 53-year collection period, VicVaDis contains a mean average of 2.5 documents per year. Our combined collection phases retrieved a peak of 11 documents published in 1889, but no texts at all for 1857–1859, 1862–1865, 1872, 1875, 1877, and 1900. As this suggests, the diachronic representation is skewed towards the latter years.

4. Case study: *Vaccination* in VicVadis

In making this corpus available as is to the scholarly community, we are aware that not all academics with an interest in its general contents will be corpus linguists, or even non-corpus linguists. Indeed, we hope that this dataset will be of interest to scholars in the social sciences and humanities, such as historians, sociologists, and political scientists, and also to researchers and practitioners in clinical, medical, and general healthcare fields. With this in mind, below, we describe an exploratory investigation of VicVaDis. Our purpose here is to exemplify accessible ways in which the data

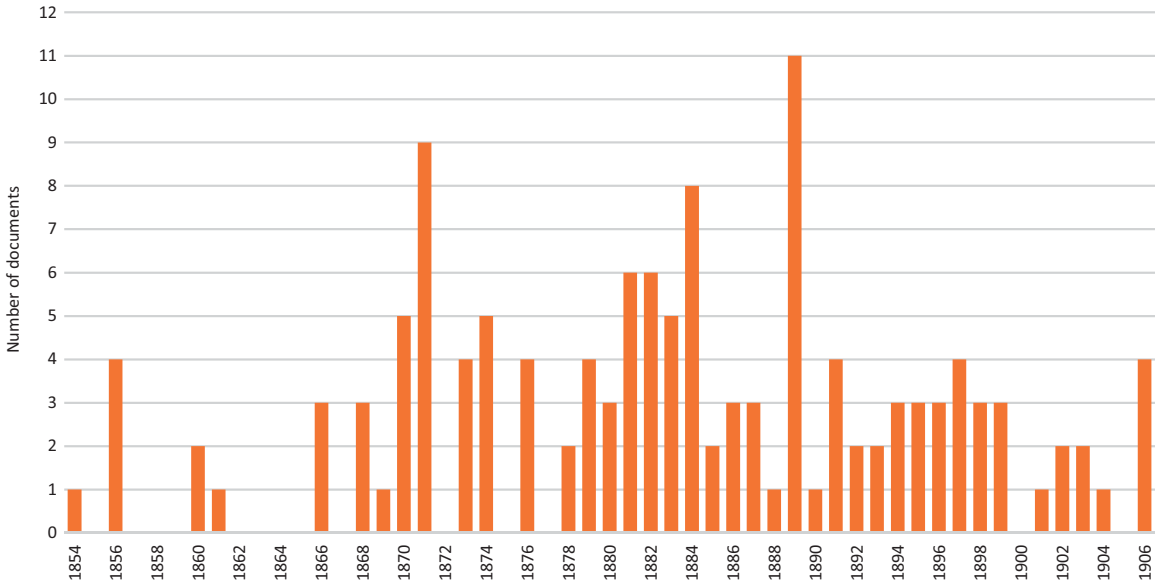


Figure 2. Chronological dispersion of texts in VicVaDis.

can be usefully interrogated that do not require either extensive expertise in corpus linguistics or access to software that presents financial or technical barriers. To achieve this, we chose a simple aim: how could we identify some of the arguments used against vaccination in our corpus? Our goal in doing so is to demonstrate what this corpus and accessible corpus tools can contribute to existing understandings of historical reasons for vaccine hesitancy. Accordingly, for this investigation, we used the freeware, AntConc to explore our VicVaDis corpus (Anthony 2022).

As is typical of many corpus investigations, we began by creating a wordlist (see Table 1) from the VicVaDis corpus. Wordlists simply count all unique words in the corpus and rank them by frequency or alphabetically. When we disregarded closed-class (grammatical) words such as *the*, *of*, *to*, *and*, *in*, and so forth, *vaccination* was the most frequent open-class (lexical) word with 31,734 occurrences.

However, such results must be taken in context. It is worth recollecting that we began compiling our corpus by using *vaccination* as a search term, so this is not an unexpected result. At the same time, that does not make it an uninsightful avenue for investigation. We therefore took *vaccination* as our first avenue of inquiry, and queried which words it most frequently co-occurred (collocated) with. We limited our collocational span to a five-word-window on either side of the word *vaccination* (−5/+5) and we report the top ten open-class results in Table 2.

Again, to keep this investigation accessible to non-specialists, and following established practice in corpus linguistics (e.g. Kilgarriff et al. 2014), we focused our attention on the top collocate, *compulsory*, and analysed a randomized sample of 500 concordance lines out of the total of 2,223. Concordance lines are instances of our search word *vaccination* and its top collocate, *compulsory*, situated within their wider linguistic context. Given the direction of this investigation, it is worth explicitly stating here that the purpose of our analysis is not to critique the arguments in our data, but simply to identify what those arguments are, what we can learn from them and, where possible, how they compare to current anti-vaccination arguments.

Table 1. Most frequent lexical words in VicVaDis, ordered by raw frequency.

Rank	Type	Raw frequency	Normalized frequency per million words
10	**vaccination**	**31,734**	**9,095.55**
19	smallpox	21,874	6,269.492
40	dr	11,186	3,206.114
45	mr	9,608	2,753.83
48	vaccinated	8,876	2,544.025
50	disease	8,592	2,462.626
53	medical	7,793	2,233.618
57	years	7,150	2,049.322
63	cases	6,258	1,793.658
69	jenner	5,345	1,531.976

4.1 Compulsory vaccination

The concordance lines for *compulsory* as a collocate of *vaccination* reveal several anti-vaccination arguments, notably, that compulsory vaccination was ineffective,

Table 2. Top 10 open-class collocates of 'vaccination' in VicVaDis with a −5/+5 window, ordered by log likelihood.

Rank	Collocate	FreqLR	FreqL	FreqR	Likelihood	Effect
1	**compulsory**	**2,223**	**1,839**	**384**	**5,583.747**	**3.032**
4	after	1,350	1,132	218	1,255.124	1.639
5	question	1,048	317	731	1,181.757	1.839
6	inquirer	379	31	348	1,053.572	3.243
8	anti	769	676	93	991.075	1.994
9	acts	453	43	410	947.557	2.697
10	against	1,055	722	333	846.51	1.504
11	act	694	99	595	749.524	1.793
13	league	251	16	235	532.783	2.723
14	tracts	184	19	165	474.361	3.087

dangerous, a threat to civil liberties and also, potentially, a distraction. For example, references to *compulsory vaccination laws* are included in texts that dispute the effectiveness of compulsory vaccination by citing smallpox death counts following its introduction:

> The **compulsory vaccination** laws came into operation in 1854 and you would naturally expect that there would be a continuous decrease in mortality. What are the facts? In 1858, 1861, 1864 and 1867, the deaths from smallpox were 6,460, 1,320, 7,684 and 2,115 respectively. (Pickering, 1871, *Vaccination: a letter in reply to an article in the "Leeds Mercury"*)

It may be useful to recap here that the seven Vaccination Acts were passed in 1840, 1853, 1867, 1871, 1874, 1898, and 1907. Given that Pickering's comments were published three decades after *Vaccination Act 1* and in the same year as *Vaccination Act 4*, it would be reasonable to assume that he was addressing *Vaccination Acts 1, 2,* and *3* which had mandated vaccines for children under three months, introduced fines for non-compliance, and then enabled repeated fines. His position, then, seems to be that the smallpox vaccine was not absolutely effective since vaccinated people were still dying from the disease, and that this therefore invalidated the arguments for mandating the smallpox vaccine. There are modern manifestations of this argument that if a given vaccine, such as the human papillomavirus (HPV) vaccine, does not guarantee protection from a disease and all its permutations either in the individual or across the population, then it is ineffectual and thus pointless. For example, in a systematic review of 103 studies of vaccine hesitancy around the HPV vaccine, Karaphillakis et al. (2019) found that an average of 34 per cent of hesitant participants across the studies cited concerns about efficacy, and 41 per cent about duration of the vaccine. This was the fourth most significant concern, following insufficient information, safety, and concerns about trust, and a more significant concern than the other six factors identified by the review.

A further examination of concordance lines where *compulsory vaccination* was followed by the third person singular verb *is* (seventy-seven occurrences) provides evidence of legal concerns, and particularly the threat to individual liberties. Examples of phrases that follow 'compulsory vaccination is' are provided in Table 3.

In some cases, this threat to civil liberties is expressed through extended metaphors. In the extract below, Dr Garth Wilkinson has sent a highly emotional and semantically loaded letter to the President and Members of the International Congress on Compulsory Vaccination, and in it, Liberty is personified as a victim of a violent attack from compulsory vaccination:

> A wider, and deeper, and subtler *Social* Evil than universal Compulsory Vaccination is scarcely conceivable; on the physical side, universal pollution; on the side of manhood, womanhood, and childhood, with their several dignities, it is to the extent of its reach, degradation and extinction. The cradle is born to an inmediate [*immediate*] medical hell. **Politically, *Compulsory Vaccination* is an innermost stab of Liberty which piercing its heart, will find its courage and licaven-born** [*heaven-born*] **principles and convictions in other directions an easy prey.** State medicine can do what it Ukes [*likes*] with us, if we once let it do this. (1879, LSACV, *The Vaccination Inquirer and Health Review: The Organ of the London Society*)

Moreover, just as the final example in Table 3 also demonstrates, Dr Wilkinson's letter intimates that the success of this attack also calls into question the safety of other 'heaven-born principles and convictions'. He therefore calls on a slippery slope argument that mandating vaccines will pave the way for future human rights infringements that will ultimately result in the destruction of the human race.

Again, one can draw parallels with these historical discussions of civil liberty and freedom and present-day reactions to, for instance, governments and employers who required travelers to have vaccination passports and employees in certain lines of work to have COVID-19 vaccinations. For instance, in January 2021, Quartz published an article on COVID-19 vaccines entitled, 'Do mandatory vaccines violate human rights?' and in July of the same year the BBC published an article entitled, 'Clubbing and Covid passports: "Protect vulnerable people" or "against civil liberties"?' Innumerable popular articles along these lines exist, demonstrating a persistent tension between

Table 3. 'Compulsory vaccination is' concordance lines in VicVaDis.

File name	*Compulsory vaccination is*. . .
1873_2_pickering_ed_the_antivaccinator.txt	a great infringement on that freedom which every man has a right to enjoy.
1879_1b_LSACV_Vacc_Inquirer.txt	a disgrace to our jurisprudence, and a shameful intrusion upon the rights of personal liberty.
1879_2b_Wilkinson_Vacc_Tracts.txt	the largest infringement of that freedom ever yet exercised.
1906_4_Furnival_Prof_Op_v2.txt	almost as disgusting as the forced creed practised by the Inquisition in mediaeval times.
1869_1_sexton_vacc_useless_injurious.txt	therefore a tyranny that everyone should strenuously resist.
1881_3b_LSACV_Vacc_Inqu_v3.txt	a system of tyranny and torhi/re [*torture*]; I use the word advisedly.
1884_1b_White_Playfair_Disposed.txt	to surrender the cardinal principle of civil and religious liberty, and to establish a precedent for the exercise of any form of tyranny.

population-level programmes imposed top-down by governments, and individual-level resistance. Giubilini (2021) explored the ethical tensions of using coercive measures to attempt to achieve herd immunity during the COVID-19 pandemic. Jecker (2022) examines the issues around vaccine passports and unequal access to health care, noting a lack of consensus both among the public and between experts.

4.2 Comparing corpora

Word lists, collocates, and concordance lines are ways into the analysis of a single corpus. To complement these, the corpus of interest can be compared and contrasted with a second corpus using the approach known as keyword analysis. This can be used to obtain a sense of the 'aboutness' of a corpus (Phillips 1989). The keyword procedure involves comparing the corpus of interest, often termed the *study corpus* or *target corpus*—in our case, VicVaDis—with a reference corpus. The purpose here is to establish which words are unusually frequent in our dataset versus a benchmark of more general language. To achieve this, keyness tools compare the relative frequencies of items (e.g. words) in the two corpora, and provide rank-ordered lists of *key*—that is, overused—items in the target corpus as compared with the reference corpus. The reference corpus is traditionally larger than the target corpus, but this is not essential (Culpeper and Demmen 2015).

To allow for maximum replicability, we searched for a suitable publicly available reference corpus, and identified *The Corpus of Late Modern English Texts v3.1* (CLMET3.1), compiled by (Diller et al. 2011).[6] CLMET3.1 contains 34 million tokens, covers the period 1790–1920, and is classified by genre. We selected only files with the following characteristics: written between 1850 and 1907, and labelled for the genres 'treatise' and 'letters'. We excluded the genres 'narrative fiction', 'narrative non-fiction', 'drama' and 'other', which contained travel writing and an extract from the magazine *Punch*. This resulted in a reference corpus, VicRef, which contains 1,947,789 tokens, of comparable dates and very approximately comparable genres to VicVaDis. We then loaded this into AntConc and generated a keyword list. Keywords ranked by likelihood establish which words are unusually frequent in the target corpus versus the reference corpus, or in our case, in VicVaDis when compared with VicRef (Table 4).

Within the top twenty-five keywords (see Table 4), we took our investigation further by considering a semantic grouping of nouns that indicate a variety of alleged vaccination-related harms, including *death*, *deaths*, *disease*, and *diseases*.

4.3 Death(s)

When we investigate these words in context, we find that *death* and *deaths* are, unsurprisingly, frequently discussed in relation to smallpox. Out of 7,207 concordance lines containing either *death* or *deaths*, we find that 861 (11.9 per cent) include examples of phrases, such as *death(s)* [*from/by/of/after*] *smallpox* or *death(s) [rate from/occurred from/caused by] smallpox*. However, interestingly, we also find that 340 (4.7 per cent) include phrases, such as *death(s)* [*from/by/of/after*] *vaccination(s)* or *death(s) [resulting from/caused by/among the/due to/occur from] vaccination(s)*, and within these discussions of deaths caused by vaccines, we find further interesting nuances (Table 5).

Within the concordance lines displayed in Table 5, we find that two include arguments that vaccine-caused deaths are being covered up. This is further demonstrated by discussion about death certificates failing to record vaccination as the underlying cause:

> In October 1876 an official inquiry was made concerning the illnesses through vaccination of sixteen children in the Misterton district of the Gainsborough Union, of which six proved fatal, but no mention was made of vaccination in any of the **death** certificates. Of the four deaths at Norwich, the subject also of an official inquiry in 1882, only one was certified as being due to vaccination. It

Table 4. Top twenty-five keywords from VicVaDis when compared with VicRef, ordered by keyness (likelihood).

Rank	Type	Raw frequency: VicVaDis	Raw frequency: VicRef	Normalized frequency per million words: VicVaDis	Normalized frequency per million words: VicRef	Keyness (Likelihood)	Keyness (Effect)
1	vaccination	31,734	4	9,095.55	2.005	28736.667	0.018
2	smallpox	21,874	4	6,269.492	2.005	19765.969	0.012
3	vaccinated	8,876	3	2,544.025	1.504	7988.536	0.005
4	**disease**	**8,592**	**116**	**2,462.626**	**58.142**	**6780.952**	**0.005**
5	dr	11,186	636	3,206.114	318.781	6459.756	0.006
6	medical	7,793	64	2,233.618	32.079	6441.045	0.004
7	jenner	5,345	0	1,531.976	0	4837.453	0.003
8	cowpox	4,687	0	1,343.381	0	4241.613	0.003
9	mr	9,608	1140	2,753.83	571.399	3731.588	0.005
10	lymph	3,586	5	1,027.814	2.506	3179.169	0.002
11	was	29,005	8671	8,313.368	4346.144	3141.183	0.016
12	vaccine	3,517	8	1,008.037	4.01	3085.139	0.002
13	inoculation	3,416	3	979.089	1.504	3048.773	0.002
14	**deaths**	**3,441**	**28**	**986.254**	**14.034**	**2844.497**	**0.002**
15	mortality	3,373	34	966.764	17.042	2739.78	0.002
16	compulsory	2,989	14	856.703	7.017	2554.487	0.002
17	epidemic	2,867	9	821.735	4.511	2490.431	0.002
18	years	7,150	997	2,049.322	499.724	2430.964	0.004
19	**diseases**	**3,059**	**40**	**876.766**	**20.049**	**2421.166**	**0.002**
20	unvaccinated	2,184	0	625.975	0	1975.893	0.001
21	cases	6,258	962	1,793.658	482.181	1940.183	0.004
22	cannot	2,116	0	606.485	0	1914.357	0.001
23	london	4,198	443	1,203.224	222.044	1770.514	0.002
24	hospital	2,276	36	652.344	18.044	1760.796	0.001
25	**death**	**3,739**	**335**	**1,071.666**	**1,67.911**	**1744.884**	**0.002**

Table 5. Five concordance lines of death* from VicVaDis.

Text file	L	Node	R
1895_5b_Hutton_Vacc_Q_v3.txt	Official denials of	deaths	from vaccination accepted
1882_2b_LSACV_Vacc_Inqu_v4.txt	in drawing attention to the cases of injury and	death	from vaccination at Norwich, and compelling
1856_1a_Gibbs_Comp_Vacc_Brief_Con.txt	occurring amongst the vaccinated, and cases of	death	from vaccination itselfarefrequent [*itself are frequent*]
1889_8_3auth_Notes_Vacc.txt	For every	death	from vaccination there are many cases of lifelong injury
1887_4_LSACV_Lead_Args_v_Comp_Vacc.txt	Officer of the Aston Union, advised in cases of	death	from vaccination to omit all mention of it from

> appeared that nine children were vaccinated in June by Dr. Guy, the public vaccinator; of these four were dead of erysipelas within three weeks of the operation, and five were seriously ill from constitutional disease. (Tebb, 1889, *What is the Truth about Vaccination?*)

This is another instance with parallels to modern discourses around vaccine-caused deaths being covered up by medical practitioners, government officials, and pharmaceutical industries. For instance, Jones et al. (2023) found that many vaccine-hesitant individuals in the UK believed both that deaths caused by the COVID-19 vaccine were concealed, and that the official numbers of deaths caused by COVID-19 itself were inflated.

More generally, the concordance lines of *death(s)* linked to vaccinations reveal stories about specific individuals, such as the following example where the death of a child is compounded by alleged fears about the consequences of mentioning to the medical practitioner that vaccination is a possible cause:

> One medical witness, after describing the **death** of a child from Vaccination, said - "He asked the

> mother what the Surgeon who performed the operation said about it. The woman's answer was - 'I dare not mention Vaccination to him, he is very cross if I do.'" (Strickland Constable, 1871, *Letter on vaccination to a medical practitioner*)

4.4 Disease(s)

The VicVaDis corpus contains 12,078 hits for *disease(s)*, and when we look at this keyword in context, we find that it can either refer to smallpox itself or to another illness. However, it does not simply refer to *the smallpox disease*, for instance, in neutral terms. Instead, this keyword regularly forms part of arguments against the effectiveness of vaccination when attempting to tackle the disease:

> There appears to be no positive security against the **disease**, either by vaccination or by smallpox inoculation, and I have seen several cases where the patients have caught smallpox twice, and have each time been very severely marked; and in two instances have died of the second attack of smallpox. (Pearce, 1868, *Vaccination: its tested effects on health, mortality, and population. An essay*)

As we mentioned earlier, there are some modern arguments that if a vaccine is not absolutely effective, it is not worthwhile (Karaphillakis et al. 2019). Further, even where *disease* does not refer to smallpox, it tends to refer to harm caused by vaccination:

> For every child that dies from smallpox, forty die from **diseases** induced by Vaccination. (LSACV, 1879, *The Vaccination Inquirer and Health Review: The Organ of the London Society*)

But what are these other diseases? To answer this question, it was necessary to step outside of a purely data-driven approach and adapt to the query at hand. Our additional purpose here is to demonstrate to non-corpus linguists that it can be as illuminating to pursue intuitive guesses and hypotheses and to explore a corpus based on our own questions and interests, and that we need not be constrained to only investigating results that are especially frequent in some way. Based on our understanding that clinical Victorian vaccine practices were radically different to those of the modern day, we therefore established potential search terms around the risks associated with the vaccine being an unintended vehicle for other diseases and harms. On that basis, we searched the corpus for *transmi**—a term that would capture *transmit*, *transmits*, *transmission*, *transmitting*, and so forth. Across the 500 hits, the concordance lines revealed an extensive list of diseases alleged to have been transmitted along with the smallpox vaccination, including *erysipelas*, *leprosy*, *suppurative meningitis*, *mesenteric disease*, *blood poisoning*, *syphilis*, *ulceration of vaccine vesicles*, and so forth. Similarly, a search for *cause**—which captures *cause*, *causes*, and *caused*—returned concordance lines with many more diseases and ailments implicated as secondary consequences of the vaccine, including *cellulitis*, *convulsions*, *eczema*, *erythema*, *gangrene*, *gangrenous eruption*, *lupus*, *pneumonia*, *puerperal fever*, *prurigo*, *pyaemia*, *scrofula*, *septicaemia*, *septic poisoning*, *syphilisation*, *syphilitic infection*, *tetanus*, and others:

> Thoughtful dentists suggest Vaccination as a probable cause of **the early decay of the teeth** in this age. (Wilkinson, 1871, *Small-Pox and Vaccination*)

> The transmission from parent to offspring of **an enfeebled constitution**, the result of vaccination. (Pearce, 1868, *Vaccination: its tested effects on health, mortality, and population. An essay*)

These discourses that causatively link modern vaccines (relatively speaking) to possible harms survive to this day, and might be most famously exemplified by Wakefield et al.'s (1998) subsequently retracted paper, which drew connections between the mumps, measles, and rubella (MMR) vaccination, the digestive disease colitis, and autism.

4.5 Old fears, new standards

As we suggested in the opening of this article, perhaps the most intriguing aspect of VicVaDis is that it demonstrates with remarkable clarity that the modern fears around new vaccines, such as those developed in light of COVID-19, are in fact not modern at all. Each new vaccine—smallpox, HPV, MMR, COVID-19—may have its unique components and concerns, but the genesis of the fears themselves appears to remain relatively stable over the centuries. These data suggest that we continue to struggle with how best to assess the risks posed by the diseases themselves versus those posed by the vaccines, how to protect our children when we are required to make decisions on their behalf that have potentially severe or even fatal outcomes, how to protect our rights to make those decisions in the face of contradictory advice or mandates, where and how to draw causative links in a febrile medical arena clouded with doubts and loudly competing voices, and so forth.

Historically, of course, such fears and hesitance were by no means groundless. The scientific understanding of diseases and the vaccination practices of the Victorian era stand in stark contrast to even the most basic medical and ethical requirements of the modern day, but similarly, in only a few decades, our current

standards and norms will seem equally outdated and even abhorrent. Medicine and clinical practice will continue to swiftly revolutionize, whereas our primal human fears can endure for centuries. It is worth stressing one final time that our purpose in this article is not to determine whether those fears around vaccinations are well founded. Instead, if history is our greatest early warning system, we can draw lessons from such corpora and our past concerns to better inform ourselves about how we are likely to respond to new developments in the future.

5. Conclusion

In this article, we introduced the VicVaDis corpus, and then presented an exploratory analysis using AntConc to look at wordlists, concordance lines, collocates, and keywords. Within this, we only sought to look at one aspect of VicVaDis—some of the arguments used against vaccinations, and to achieve this we examined the top collocates and keywords in our corpus. Even within this, we contemplated only a minority of the overall results. Many more avenues were open to investigation that there was simply insufficient room to fully explore. Future researchers might, for instance, compare modern corpora of ephemeral anti-vaccine content such as blog posts with VicVaDis to determine how much similarity or difference exists between them. Alternatively, it would be possible to investigate how discourse around sexual promiscuity in a corpus of HPV vaccine discussions compares to the arguments found in VicVaDis that poor personal hygiene amongst the working class was primarily responsible for smallpox. One could also contrast discussions of COVID-19 vaccine efficacy with the arguments of the efficacy of the smallpox vaccine in our corpus.

Whatever the choice of avenue, we hope that this article provides an interesting beginning not just for corpus linguists, but also for those working in related humanities, social science and medical fields who may wish to explore this dataset and approach for themselves. To facilitate this, we are making VicVaDis available for anyone to use as is, along with other relevant and useful resources on its accompanying webpages.[7]

Author contributions

Claire Hardaker (Conceptualization, Funding acquisition, Investigation, Methodology, Writing—original draft), Alice Deignan (Conceptualization, Investigation, Methodology, Writing—original draft), Elena Semino (Conceptualization, Funding acquisition, Investigation, Project administration, Writing—original draft), Tara Coltman-Patel (Conceptualization, Data curation, Writing—review and editing), William Dance (Conceptualization, Data curation, Writing—review and editing), Zsófia Demjén (Conceptualization, Funding acquisition, Writing—review and editing), Chris Sanderson (Data curation, Methodology, Writing—review and editing), Derek Gatherer (Conceptualization, Data curation, Funding acquisition, Investigation, Methodology, Resources, Writing—original draft)

Supplementary data

Supplementary data is available at *DSH* online.

Conflict of interest statement. None declared.

Funding

This study was funded by the Economic and Social Research Council, part of UK Research and Innovation (grant no. ES/V000926/1).

Notes

1. A Royal Commission is a major public enquiry.
2. https://wellcomecollection.org/, https://www.gutenberg.org/, and https://oll.libertyfund.org/, respectively.
3. https://www.medicalheritage.org, https://www.bl.uk/, https://www.jstor.org/, and https://archive.org/, respectively. These resources linked to documents in the following further archives: the U.S. National Library of Medicine, Bristol Selected Pamphlets, the London School of Economics Selected Pamphlets, the Francis A. Countway Library of Medicine, the Harold B. Lee Library, the London School of Hygiene & Tropical Medicine Library & Archives Service, the Royal College of Physicians in Edinburgh, the Royal College of Surgeons of England and the online collections of Harvard University, Oxford University, Saint Mary's College of California, University of California, University of Glasgow, University of Leeds, and the Cushing/Whitney Medical Library at Yale University. Additional few documents were uploads by private individuals to the Internet Archive.
4. Vaccination laws were passed later for Scotland and the island of Ireland.
5. Declared authorship is not automatically the same as executive authorship. For instance, a monograph's *declared* author may be the LSACV, and we may never be able to divine the identity of the *executive* author(s)—the person or people who actually wrote the text.
6. http://fedora.clarin-d.uni-saarland.de/clmet/clmet.html
7. The texts in VicVaDis are all drawn from the public domain and are all outside of copyright. Information on downloading the corpus is available and updated as necessary on the project website: https://www.lancaster.ac.uk/vaccination-discourse/.

References

Arnold, W., and Arnold, C. (2022) 'Medicine in the 'Monthly Review': Revealing Public Medical Discourse through Topic Modelling', *Digital Scholarship in the Humanities*, **37**: 611–29.

Altick, R. (1957) *The English Common Reader: A Social History of the Mass Reading Public*. Columbus, OH: Ohio State University Press.

Anthony, L. (2022) *AntConc (Version 4.2.0) [Computer Software]*. Tokyo, Japan: Waseda University.

Atkins, S., Clear, J., and Ostler, N., (1992) 'Corpus Design and Criteria', *Literary and Linguistic Computing*, **7**: 1–16.

Biber, D. (1993) 'Representativeness in Corpus Design', *Literary and Linguistic Computing* **8**: 243–57.

Bonanni. P., and Santos, J. (2011) 'Vaccine Evolution', *Understanding Modern Vaccines: Perspectives in Vaccinology*, **1**: 1–23.

Callender, D. (2016) 'Vaccine Hesitancy: More than a Movement', *Human Vaccines and Immunotherapeutics*, **12**: 2462–68.

Charlton, C. (1983) 'The Fight Against Vaccination: The Leicester Demonstration of 1885', *Local Population Studies*, **30**: 60–66.

Coltman-Patel, T. *et al.* (2022) 'Am I being Unreasonable to Vaccinate my Kids against My Ex's Wishes?'–A Corpus Linguistic Exploration of Conflict in Vaccination Discussions on Mumsnet Talk's AIBU Forum', *Discourse, Context & Media*, **48**: 1–9.

Culpeper, J., and Demmen, J. (2015) 'Keywords'. In: Biber, D. and Reppen, R. (eds) *The Cambridge Handbook of English Corpus Linguistics*, pp. 90–105. Cambridge: Cambridge University Press.

Diller, H., De Smet, H., and Tyrrkö, J. (2011) 'A European database of descriptors of english electronic texts', *The European English Messenger*, **19**: 21–35.

Durbach, N. (2000) 'They Might as Well Brand Us: Working Class Resistance to Compulsory Vaccination in Victorian England', *The Society for the Social History of Medicine*, **13**: 45–62.

Durbach, N. (2005) *Bodily Matters. The Anti-Vaccination Movement in England, 1853-1907*. Durham, NC: Duke University Press.

Fajri Nuwarda, R. *et al.* (2022) 'Vaccine Hesitancy: Contemporary Issues and Historical Background', *Vaccines*, **10**: 1–21.

Giubilini, A. (2021) 'Vaccination Ethics', *British Medical Bulletin*, **137**: 4–12.

Hobbs, A. (2018) *A Fleet Street in every town: The provincial press in England 1855-1900*. Cambridge: Open Books.

Humphries, B. (2011) Nineteenth Century Pamphlets Online. *The Ephemerist*, 153.

Jecker, N. (2022) 'Vaccine Passports and Health Disparities: A Perillous Journey', *BMJ: Journal of Medical Ethics*, **48**: 957–60.

Jones, L. *et al.* (2023) 'Understanding the Public's Attitudes towards Covid-19 Vaccinations in Nottinghamshore, United Kingdon: Qualitative Social Media Analysis', *Journal of Medical Internet Research*, **25**.

Karaphillakis, E. *et al.* (2019) 'HPV Vaccination in a Context of Mistrust and Uncertainty: A Systematic Literature Review of Determinants of HPV Vaccine Hesitancy in Europe', *Human Vaccines Immunotherapeutics*, **15**: 1615–27.

Kilgarriff, A. *et al.* (2014) 'The Sketch Engine: Ten Years On', *Lexicography*, **1**: 7–36.

Klondrlik, K. E. (2020) 'Conscientious Objection to Vaccination and Failure to Solidify Professional Identity in Late Victorian Socio-medical Journals', *Victorian Periodicals Review*, **53**: 338–71.

Lloyd, A. (2007) *Education, Literacy and the Reading Public. British Library Newspapers*. Detroit: Gale.

McEnery, T., and Hardie, A. (2011) *Corpus Linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press.

Phillips, M. (1989) *Lexical Structure of Text [Discourse Analysis Monograph 12]*. Birmingham, UK: University of Birmingham.

Puri, N. *et al.* (2020) 'Social Media and Vaccine Hesitancy: New Updates for the Era of Covid-19 and Globalized Infectious Diseases', *Human Vaccines and Immunotherapeutics*, **16**: 2586–2593.

Roberts, L. (2020) 'Measles is on the Rise, and Covid-19 could Make it Worse', *Nature*, **580**: 447–48.

Stewart, A. J., and Devlin, P. M. (2006) 'The History of the Smallpox Vaccine', *Journal of Infection*, **52**: 329–34.

Swales, J. D. (1992) 'The Leicester Anti-vaccination Movement', *Lancet*, **340**: 1019–21.

Tafuri, S. *et al.* (2014) 'Addressing the Anti-vaccination Movement and the Role of HCWs', *Vaccine*, **32**: 4860–65.

Trogen, B., and Pirofski, L. (2021) 'Understanding Vaccine Hesitancy in COVID-19', *Med Commentary*, **2**: 498–500.

Wakefield, A. J. *et al.* (1998) 'Ileal-lymphoid-nodular Hyperplasia, Non-specific Colitis, and Pervasive Developmental Disorder in Children', *Lancet*, **351**: 637–41.

Weber, T. P. (2010) 'Alfred Russel Wallace and the Antivaccination Movement in Victorian England', *Emerging Infectious Diseases*, **16**: 664–68.

World Health Organization (2019) 'Top Ten Threats to Global Health in 2019', https://www.who.int/news-room/spotlight/ten-threats-to-global-health-in-2019, accessed 9 Dec. 2022.

Williams, N. (1994) 'The Implementation of Compulsory Health Legislation: Infant Smallpox Vaccination in England and Wales, 1840-1890', *Journal of Historical Geography*, **20**: 396–412.

Williamson, S. (2007) *The Vaccination Controversy. The Rise, Reign and Fall of Compulsory Vaccination for Smallpox*. Liverpool: Liverpool University Press.