Hunting for Previous Coronavirus Pandemics Using Corpus Linguistic Analysis of 19th Century British Newspapers

Rosita Solovejute & Derek Gatherer*

Division of Biomedical & Life Sciences

Faculty of Health & Medicine

Lancaster University

LA1 4YW, UK

*Corresponding author

E-mail: d.gatherer@lancaster.ac.uk

 doi:10.20944/preprints202012.0817.v2

Abstract

COVID-19 is the first known coronavirus pandemic. Nevertheless, the seasonal circulation of the four milder coronaviruses of humans – OC43, NL63, 229E and HKU1 – raises the possibility that these viruses are the descendants of more ancient coronavirus pandemics. This proposal arises by analogy to the observed descent of seasonal influenza subtypes H2N2 (now extinct), H3N2 and H1H1 from the pandemic strains of 1957, 1968 and 2009, respectively. Recent historical revisionist speculation has focussed on the influenza pandemic of 1889-1892, based on molecular phylogenetic reconstructions that show the emergence of human coronavirus OC43 around that time, probably by zoonosis from cattle. If the "Russian influenza", as *The Times* named it in early 1890, was not influenza but caused by a coronavirus, the origins of the other three milder human coronaviruses may also have left a residue of clinical evidence in the 19th century medical literature and popular press. In this paper, we search digitised 19th century British newspapers for evidence of previously unsuspected coronavirus pandemics. We conclude that there is little or no corpus linguistic signal in the UK national press for large-scale outbreaks of unidentified respiratory disease for the period 1785 to 1890.

Keywords

COVID-19; OC43; coronavirus; pandemic; influenza; historical revisionism; corpus linguistics

Introduction

The article "Did a coronavirus cause the pandemic that killed Queen Victoria's heir?" by journalist Robin McKie [1], which appeared in UK Sunday newspaper *The Observer* on 31st May 2020, presents an intriguing piece of historical revisionism. Ever since the late 19th century, it has been generally accepted that the pandemic of respiratory illness that occurred between 1889 and 1892, was caused by influenza. *The Times* newspaper of London first referred to it as "*the Russian influenza*" on 6th

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 December 2022 doi:10.20944/preprints202012.0817.v2

January 1890 and the excess death toll by the end of 1892 in the United Kingdom was 110,000 [2], slightly under half the number of deaths that occurred in the UK during the so-called "Spanish flu" pandemic of 1918-1920 [3]. Samples derived from the graves of pandemic victims from 1918 prove beyond doubt that they died from influenza type A subtype H1N1 [4]. However, similar samples from the early 1890s have so far been unobtainable, leaving open the possibility that the causative agent of that previous pandemic was something other than an influenza virus.

McKie's article draws attention to one of the most famous victims of *"the Russian influenza*" - the Duke of Clarence, second in line to the throne, whose illness was first reported by *The Times* on 11th January 1892:

> *We regret to announce that the Duke of Clarence and Avondale, who is with the Prince and Princess of Wales at Sandringham, is suffering from a severe attack of influenza, accompanied by pneumonia. A telegram last evening from Sandringham states that his Royal Highness's strength is well maintained. Dr. Laking* [On Sir Francis Henry Laking's (1847-1914) considerable posthumous notoriety, see: 5] *has been at Sandringham since Saturday. All the Duke's engagements have, of course, for the present been cancelled. His Royal Highness's 28th birthday was celebrated at Sandringham by a dinner party on Friday.*

The following morning *The Times* observed that *"a large number of the Prince of Wales's household are down with influenza"*. Reports on the health of the Duke were initially optimistic: his temperature had returned to normal, he had slept. However, by the following day, the descriptions of the Duke's condition had become a little more ambiguous. A drop in the stock market was blamed on nervousness concerning the effect the Duke's illness might have on the line of succession. The next day's early editions described the Duke as *"almost in the jaws of death"*. By the time those newspapers were being read, he had already died.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 December 2022 doi:10.20944/preprints202012.0817.v2

Retrospective diagnosis is always difficult, especially at historical distances of time, and there are many pathogens that could have produced the pneumonia that killed the Duke. However, McKie's suggestion of a coronavirus rather than influenza is not without some support in the scientific literature. Vijgen et al. [6] produced the first complete genome of human coronavirus OC43 (family *Coronaviridae*; sub-family *Orthocoronavirinae*; genus *Betacoronavirus*; sub-genus *Embecovirus*; species name *Betacoronavirus 1*) and dated its common ancestor with its nearest relative - bovine coronavirus (same taxonomic hierarchy) - at 1873 (95% confidence interval [CI] 1815-1899), using a correlation of root-to-tip genetic distances and root-to-tip time intervals on a maximum likelihood phylogenetic tree. Two variants of Bayesian phylogenetic estimation also produced divergence dates from bovine coronavirus at 1890 (95% CI 1859-1912; coalescent model) and 1893 (95% CI 1866-1918; exponential growth model). Additionally, Vijgen et al. pointed out that the period 1870-1890 saw mass culling of cattle due to an outbreak of infectious respiratory disease. If that disease of cattle was caused by bovine coronavirus, Vijgen et al. speculated, then such culling provides an obvious candidate process for effective zoonosis of the common ancestor of bovine coronavirus and human OC43, from cattle to humans. A final piece of circumstantial evidence was provided by the observation that *"the Russian influenza"* involved frequent neurological complications. This is not typical of influenza infection but is found in some OC43 patients [7, 8].

McKie's topical revival in 2020 of the 2005 Vijgen et al. hypothesis, has led to further investigations in this area. Erkoreka et al. [9] focus on various clinical and epidemiological parameters of *"the Russian influenza"* which they conclude are more similar to those of COVID-19 than to any known influenza pandemic. Erkoreka et al. firmly align themselves on the side of a coronavirus as the causative agent of *"the Russian influenza"* pandemic. Brussow and Brussow [10] and Berche [11] perform similar analyses and, although also sympathetic to that theory, are slightly more agnostic in terms of conclusions. Forni et al. [12] update the molecular phylogenetic analysis to incorporate the genome sequences of the many strains of both OC43 and bovine coronavirus that have become available since 2005. By contrast, those authors produce a mean divergence date of the two viruses that is slightly

 doi:10.20944/preprints202012.0817.v2

too late, at 1923 (95% CI 1872–1967), to permit OC43 to be the perfect candidate pathogen for *"the Russian influenza"*. Nevertheless, they note that 1889 does lie within their 95% confidence interval, while providing a cautionary discussion on technical issues of dating the divergence.

Defenders of the accepted theory that *"the Russian influenza"* was influenza will, at this juncture, point to the evidence of seroarchaeology – the detection of antibodies against influenza in historically donated sera. Worobey et al. [13] produced a comprehensive attempt to synthesize the findings of seroarchaeology with more recent advances in influenza genome sequencing, concluding that the evidence points towards influenza type A subtype H3N8 as the cause of *"the Russian influenza"*. This also explains the higher excess mortality from 1918's "Spanish flu" among those born in the late 1880s, since their principal exposure to influenza, especially during the crucial early childhood period, would have been to H3N8, giving them poor protection on the arrival of H1N1 in 1918. Older adults, by contrast, may have been exposed to seasonal influenzas circulating pre-1889, possibly containing H1, N1 or both, and thus have been better prepared for "Spanish flu" when it arrived in 1918. Furthermore, those survivors of *"the Russian influenza"* who lived long enough to see the "Hong Kong flu" pandemic of 1968, caused by influenza type A subtype H3N2, had a lower mortality rate than younger individuals, providing a further piece of evidence for exposure to influenza H3 antigens in the 1890s. Some preliminary results of palaeoserology – the detection of antibodies in cadavers – have detected evidence of coronavirus exposure in individuals born between 1864 and 1894, but it is not possible to identify the exact virus to which they were exposed, nor the date of their exposure [14].

Therefore, at present, the evidence on both sides of the question remains circumstantial. In the absence of any clinical samples from *"the Russian influenza"*, a decision between one theory or the other depends largely on how one weights the importance of the various circumstances. Even if one rejects the candidacy of OC43 as a potential aetiological agent of *"the Russian influenza"* pandemic, the possibility remains that the four seasonal coronaviruses of humans are indeed the descendants of more historically distant unidentified pandemics. Although it is important not to draw too many

 doi:10.20944/preprints202012.0817.v2

analogies between coronaviruses (family *Coronaviridae*) and influenza viruses (family *Orthomyxoviridae*) – their common properties are virtually limited to the observations that they are respiratory, potentially fatal, moderately to highly infectious and have incubation periods of a few days – the inevitable annual return of seasonal influenzas H3N2 and H1N1, descended from the pandemics of 1968 and 2009 respectively, provides a persuasive model for the origins of the more irregular seasonal return of the four milder human coronaviruses.

If pandemics of respiratory disease occurred prior to 1890, they may have been remarked upon in the newspapers of the time. In this study, we perform a search of a digitized corpus of national and local British newspapers from the 19th century, using search terms indicative of coronavirus symptoms. We plot the occurrence of these textual elements over time and identify years when they were significantly elevated (p<0.05). For candidate dates identified using this corpus linguistic method, we review the newspaper texts in question for further contextual information. For those instances where outbreaks of respiratory disease are described, we search the medical literature around those dates for the contemporary attribution of their cause.

Methods

Search Terms

For clarity, in the ensuing descriptions, we use **bold** font for search terms and *italic* font for collocates and other quotations.

Based on clinical descriptions of COVID-19 [reviewed: 15], we selected the following search terms: 1) "**cough**", 2) "**fever**", 3) "**pneumonia**". To avoid confusion with years when influenza pandemics may have occurred, we added 4) "**influenza**" and 5) "**epidemic**". Any combination of terms 1 to 3 co-occurring with term 4 alone or terms 4 and 5 together, would be indicative of a respiratory outbreak caused by, or at the least attributed to, influenza. By contrast, any combination of terms 1 to 3 co-

 doi:10.20944/preprints202012.0817.v2

occurring with term 5 alone, or without either of terms 4 and 5, would suggest a respiratory disease that was not confidently identified as influenza at the time. This outbreak would provide a candidate coronavirus epidemic for further investigation.

Newspapers

Newspapers and years searched were as follows: *Belfast Newsletter* (1828-1900), *The Era* (1838-1900), *Glasgow Herald* (1820-1900), *Hampshire & Portsmouth Telegraph* (1799-1900), *Ipswich Journal* (1800-1900), *Liverpool Mercury* (1811-1900), *Northern Echo* (1870-1900) *Pall Mall Gazette* (1865-1900), *Reynold's Daily* (1850-1900), *Western Mail* (1869-1900) and *The Times* (1785-2009). The search in *The Times* was extended to 2009 in order to provide a comparison with the 20th century.

Searches were performed using Lancaster University's instance of the CQPweb (Corpus Query Processor) corpus analysis software [https://cqpweb.lancs.ac.uk/; 16]. CQPweb's database is populated from the newspapers listed, using optical character recognition (OCR), so for older publications in particular, some errors may be present [17].

Statistics

The occurrence of each of the five search terms was calculated per million words within the annual output of each publication, in CQPweb. This is compared to a background distribution constituting the corresponding words per million for each search term over the total year range for each newspaper. Within the annual distributions, for each search term and each newspaper, we determined the years lying in the top 1% (i.e. $p<0.05$ after application of a Bonferroni correction), following Gabrielatos et al. [18]. These are deemed to be years when that search term was in statistically significant usage above its background level for the newspaper in which it occurs. For years when search terms were significantly elevated, we also calculated collocates at range *n*. Collocates, in corpus linguistics, are other words found at statistically significant usage, over their own background levels, in a window from *n* positions to the left to *n* positions to the right of the search

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 December 2022 doi:10.20944/preprints202012.0817.v2

term. In other words, they are found in significant proximity to the search term. A default value of *n*=10 was used throughout, unless specified. Collocation analysis therefore assists in showing how a search term associates with other words within a corpus, providing information about the context in which that search term is used. CQPweb provides a log ratio method for the quantification of the strength of collocation.

Results

When each year was examined in more detail it was found that search term "**cough**" was prone to artefactual hits due to marketing campaigns for patent medicines, namely "*Du Barry's Excellent Revalenta Arabica Food*" in 1854, "*Dr Locock's Pulmonic Wafers*" in 1855, "*Tamarind Cough Emulsion*" in 1857, "*Taylor's Cough Elixir*" in 1876, "*Powell's Balsam of Aniseed*" in 1881 and 1886 and "*Keating's Cough Lozenges*" in 1860, 1879, 1880, 1886, 1887 and 1896. 1860 was a particularly intense year for patent medicine advertising with "*Gostling's Cough Balls*", "*Smith's Sedative Cough Candy*" and "*Pectoral Cough Lozenges*" also contributing to the statistical signal for "**cough**". These instances are deleted from Table 1. A "**cough**" signal from the year 1888 derived from accounts of the death of Emperor Frederick of Germany from throat cancer along with some concurrent, perhaps opportunistic, marketing of throat lozenges. This has also been deleted. Following this cleaning operation, Table 1 summarizes the years when each of the search terms achieved statistically significant usage at $p<0.05$.

Year	Influenza	Pneumonia	Cough	Fever	Epidemic	Comments
1790			1			whooping cough
1791			1			whooping cough
1792			1			whooping cough
1793			1			whooping cough
1804				1		yellow fever & typhus abroad
1817				1		typhus
1819				1		yellow fever, scarlet fever, measles & typhus
1832					1	cholera
1847				3		typhus

1848				1	1	typhus, scarlet fever
1849				1	3	cholera, typhus
1852				1		typhus, cholera, yellow fever
1853				1	2	typhus, cholera, yellow fever
1854					1	cholera
1865					2	typhus, cholera & a "Russian epidemic"
1866					2	cholera
1870				1		scarlet fever, typhoid, typhus, smallpox, diphtheria, yellow fever
1873			1			whooping cough
1876			1	1		scarlet fever, typhoid, diphtheria, yellow fever, whooping cough
1877				1	1	scarlet fever, typhoid, diphtheria, yellow fever, whooping cough
1878			1	1		scarlet fever, typhoid, diphtheria, yellow fever, whooping cough
1881				1		scarlet fever, typhoid, typhus
1884					1	cholera abroad, measles, sore throats, whooping cough, smallpox in India
1887				1		malaria, scarlet fever, typhoid, typhus, smallpox, diphtheria, measles
1890	7	1			6	Russian flu, typhoid, scarlet fever, typhus, diphtheria, cholera
1891	5			2	2	Russian flu, typhoid, scarlet fever, malaria, swine fever, typhus, smallpox, measles, yellow fever, diphtheria,
1892	9	2	1	1	9	Russian flu, typhoid, scarlet fever, malaria, swine fever, typhus, smallpox, measles, yellow fever, diphtheria, cholera, whooping cough
1893	2	1	2			Russian flu, whooping cough
1894			1	1		scarlet fever, typhoid, swine fever, whooping cough
1895	3		1	1		scarlet fever, typhoid, swine fever, influenza, whooping cough
1896			1	1	1	scarlet fever, typhoid, swine fever, whooping cough, measles, diphtheria

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 December 2022 doi:10.20944/preprints202012.0817.v2

1899		2				pneumonia (S. Africa)
1900		8	1	3		swine fever (UK), typhoid, malaria, pneumonia (all S. Africa)
1901		1		1		typhoid, malaria, pneumonia (all S. Africa)
1902		1		1		typhoid, malaria, pneumonia (all S. Africa)
1918	1	1			1	Spanish flu, typhus, cholera
1919	1	1				Spanish flu
1957	1					H2N2 pandemic
2001					1	foot-and-mouth, measles, AIDS, influenza
2005	1					bird flu H5N1
2006	1					bird flu H5N1
2009	1					swine flu H1N1

Table 1: Statistically significant occurrences of search terms. The numbers refer to the number of newspapers in which that search term was statistically significant at $p<0.05$ (after Bonferroni correction) for that year. Only years with statistically significant occurrences in at least one publication are listed. Prior to 1799 and from 1901 onwards, all numbers are 1, since only *The Times* was available to be searched for those years. Years of acknowledged influenza pandemics are shaded in red, years of cholera pandemics in green and years when these two categories overlap in blue. Some acknowledged influenza pandemic years are absent from the table (1830-31, 1833, 1968) because there are no statistically significant hits to the five search terms.

We deal with each search term in turn, listing the years when it achieved statistical significance at $p<0.05$ after Bonferroni correction [19] and reporting their scores in words/million. If that score is non-significant, it is indicated in the text as "non-sig.". All scores without this indication can be assumed to be statistically significant. For those who do not wish to read the detail for each search term on a year-by-year basis, a summary is provided at the bottom of each search term subsection.

Influenza

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 December 2022 doi:10.20944/preprints202012.0817.v2

It should be noted that statistical significance at $p<0.05$ for "**influenza**" is not achieved for any years prior to 1890. However, since McKie [1] and Vijgen et al. [6] postulate that at least one ostensible influenza pandemic may have been in reality a coronavirus pandemic, all years when the distribution of "**influenza**" showed a peak are considered, as well as the "official" influenza pandemic years specified by Worobey et al. [13], namely 1830-1833 [see also 20] and 1847-1850. The pandemic of 1830-1833 appears to have struck the UK in its later stages, in 1833, as judged by occurrences of "**influenza**" (1.94 words/million, non-sig.). All but two of the occurrences are from the 16th April to the 24th June, for instance the following upbeat message on the 1st May 1833:

> *The grippe, or influenza, coming from the east, is approaching us …. it is now said that there have been some cases here since yesterday morning. We do not fear it, however, as it is attended with little inconvenience, and no danger*.

Evidently not all were reassured, as on the 7th May, *The Times* felt the need to complain that:

> *The alarmists have begun to raise the most absurd and boundless apprehensions in the public mind on the subject of the present epidemic*.

On the 16th May, it was pointed out that things were much worse in France:

> *It is calculated that there are at this moment about 100,000 individuals in Paris who are under the influence of this strong disorder*.

Annoyingly, the weather refused to make any difference as the epidemic had:

> *… lately become very troublesome here, in spite of the truly beautiful weather and steady temperatures that we have been enjoying since the close of the last month*.

Such virulence during summer months is typical of a pandemic influenza, and was seen in the H1N1 pandemic of 2009 [21].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 December 2022 doi:10.20944/preprints202012.0817.v2

Although 1833 is a recognised influenza pandemic year, mentions of "**influenza**" peaked even higher in *The Times* in 1837 (4.85 words/million, non-sig.). However, in 1837, the distribution of occurrences differs. All but 12 mentions occur before the 30th May. Although the pattern in 1833 certainly suggests a pandemic, that in 1837 might represent simply a prolonged and severe seasonal outbreak or a "second wave" phase of the pandemic of 1833.

In 1847, "**influenza**" is mentioned at 1.14 words/million (non-sig.). Reports began in the foreign news section of *The Times* on the 2nd December 1847: *"The influenza continued to spread in an alarming manner in Paris"* and by 13th December it was referred to as *"the prevailing epidemic"* in London, especially in "*the most unhealthy spots*" such as "*East London, Poplar and Camberwell*". As Christmas approached *The Times* was also bemoaning "*the number of those who are at present incapacitated by influenza"*. The brunt of the epidemic in the UK appears to have been felt in 1848 when *The Times* has "**influenza**" at 2.72 words/million (non-sig.), principally in January and February.

Nevertheless, as indicated by the qualification "non-sig." following all of the above occurrence statistics, although we observe peaks in 1833, 1837 and 1848, at no point does "**influenza**" creep above the significance threshold. This is perhaps due to the elevation of that threshold by the extremely high number of mentions from 1890-1893 and 1895 (peaking at 21.0 words/million in 1892, although quite a lot of that coverage revolved around the unfortunate Duke of Clarence). After *"the Russian influenza"* pandemic, "**influenza**" peaks again in 1900 in *The Times* (6.87 words/million, non-sig.). 193 deaths from influenza were recorded in London in the week ending 3rd January 1900, rising to 340 in the week to 17th January, declining a month later to 63.

Summary: the statistically significant signal for "**influenza**" from 1785 to 1900 is dominated by *"the Russian influenza"* pandemic and spans the years 1890-93 inclusive plus 1895. Nevertheless, smaller, non-significant peaks are observed for the other known pandemic years of 1833 and 1848. Among accepted non-pandemic years, 1837 and 1900 have peaks and indeed 1837 is higher than 1833. After 1900, statistically significant signals are observed for the "Spanish flu" in 1918 and 1919, for the "Asian

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 December 2022 doi:10.20944/preprints202012.0817.v2

flu" in 1957, for the "bird flu" scare in 2005 and 2006 and for the H1N1 pandemic in 2009. A non-significant peak is observable for "Hong Kong flu" in 1969. We did not examine the 20th and 21st century peaks in detail as they are all explicable by influenza pandemics for which clinical samples are available and there is no doubt that the causative agent was influenza and not a coronavirus.

Pneumonia

Like "**influenza**", there are no significant signals for "**pneumonia**" prior to 1890. The first statistically significant peaks for "**pneumonia**" overlap those of *"the Russian influenza"* years 1890-1893, with the exception of 1891. A second significant peak for "**pneumonia**" is found in years 1899-1902. This generally refers to the Boer war when many deaths from pneumonia were recorded among British troops in South Africa. Overall, "**pneumonia**" presents a generally flatter distribution across the years than that of "**influenza**" and, unlike "**influenza**", there are no peaks that might be termed "non-significant yet notable". In particular, we note that there is no signal for a pneumonia epidemic independent of influenza, which might be potentially characteristic of a coronavirus.

Summary: With the exception of the Boer War years, "**pneumonia**" only occurs at statistically significant levels during recognised influenza pandemic years. The symptom of pneumonia appears to be overwhelmingly associated with the presence of influenza in the population. There is therefore no evidence of a non-influenza pneumonia outbreak in the period from 1785 to 1890, or at least no pneumonia outbreak that was not attributed to influenza during this period.

Cough

The first statistically significant years for mention of "**cough**" are from 1790-1793 in *The Times*, peaking at 10.9 words/million in 1792. These predominantly refer to whooping cough (spelled "*hooping cough*" in the 1790s), now known to be caused by *Bordatella pertussis*. In *The Times*, for the decade of the 1790s, the highest collocates at length n=1 – meaning the words most commonly found to the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 December 2022 doi:10.20944/preprints202012.0817.v2

immediate left and right of "**cough**" – are "*hooping*", "*violent*", "*dreadful*", "*continued*", "*dry*" and "*bad*". Of these "*hooping*" is the most statistically significant with a log ratio of 18.5 for the 1790s.

The same disease, now spelled in the modern manner, resurfaces as a statistically significant signal for "**cough**" in the *Hampshire & Portsmouth Telegraph* in 1873 with 6.64 words/million. The highest collocation at a wider window length *n*=10 is "*whooping*" at log ratio 17.9. Three years later in 1876, the same newspaper also displays "**cough**" at 5.79 words/million, reporting on the 15th April 1876 that the death rate in Portsmouth from "*seven principal zymotic* [i.e. infectious] *diseases*" is 8.3 per thousand over the previous 12 months – a total of 407 deaths, of which 143 are recorded as due to whooping cough. This would give whooping cough a mortality rate of 2.9 deaths per thousand per annum in the city of Portsmouth in 1876. Another two years later, in 1878, "**cough**" surfaces again in the *Hampshire & Portsmouth Telegraph* at 12.8 words/million and once more "*whooping*" is the highest collocate at log ratio 16.9.

The Times shows statistically significant signals for "**cough**" from 1892 to 1896 inclusive, partly overlapping the period of *"the Russian influenza"*. The vast majority of these are advertising for cough remedies. The remainder overwhelmingly refer to whooping cough at collocation log ratio 17.6. The same pattern is found in the *Ipswich Journal*, but restricted to 1893, and in the *Western Mail* in 1900.

Summary: once influenza pandemic years are removed, the statistically significant usage of "**cough**" refers overwhelmingly to whooping cough, a disease as well recognised then, due to its distinctive symptoms, as it is now. There is therefore no evidence of a cough-associated non-whooping cough outbreak in the period from 1785 to 1890.

Fever

The *Hampshire & Portsmouth Telegraph* has a hit for "**fever**" in 1804 at 75.4 words/million. However, this mostly refers to military situations abroad, including Savannah, Gibraltar, the Cape of Good Hope, Antigua, Santo Domingo and Dalmatia among others. The fever described is most often yellow fever,

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 December 2022 doi:10.20944/preprints202012.0817.v2

a tropical disease now known to be caused by mosquito-transmitted yellow fever virus – "*yellow*" has a collocation at log ratio 10.2. Other references to "**fever**" describe "*putrid fever*" – a lice-borne disease now known as typhus, caused by *Rickettsia prowazekii* – and, again overseas, "*jungle fever*", of less certain aetiology [some later authors used it as a synonym for dengue virus e.g. 22]. Among the fewer references to domestic fever outbreaks, on 29th October 1804, we find mention of another well recognised disease, scarlet fever, now known to be caused by *Streptococcus pyogenes*:

> *… in consequence of the prevalence of the scarlet fever at Westminster school, an extraordinary vacation, for the term of a fortnight, has taken place*.

The next significant appearance of search term "**fever**" is in the *Liverpool Mercury* in 1817 at 44.0 words/million. In this instance, the top collocate is "*typhus*" at log ratio 16.1. Most of these are cases in Ireland, with which Liverpool had strong contact links and therefore high likelihood of incoming transmission of the outbreak [see also 23]. For instance on 12th September 1817 there is a report on the:

> *… alarming progress of the typhus fever in Ireland. Not one part of the country is free from it. This wide and rapid extension of the disease has been effected by the wandering hordes of beggars who traverse the island in every direction in search of food, to prolong their miserable existence*.

In *The Times* in 1819, "**fever**" is found at 39.5 words/million. These are a mixture of descriptions of yellow fever, scarlet fever, measles and typhus, with the latter described on 23rd July 1819 as "*the present prevailing epidemic*". The spread of typhus to almost all parts of the country is shown by a report on the 17th September:

> *We regret to state that the typhus fever is rapidly spreading in the town of Montrose* [Forfarshire, far from Ireland]*; so much so, that the Chief Magistrate has judged it necessary to call a general meeting of the citizens on the subject*.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 December 2022 doi:10.20944/preprints202012.0817.v2

Search term "**fever**" reaches statistical significance again in 1847 in the *Liverpool Mercury*, *Glasgow Herald* and *Belfast Newsletter*. This was the year that typhus also made a literary appearance in Charlotte Bronte's novel *Jane Eyre*. 1847 was also an influenza pandemic year although, as mentioned above, influenza peaked the following year in the UK. The *Belfast Newsletter* has search term "**fever**" at 112 words/million with "*typhus*" again featuring as a collocate at log ratio 12.7. On 12th February 1847, it was reported that:

> *Pestilence, was now beginning to complete the work of famine -the typhus fever was raging in Ireland - and if the national generosity were not extended to that country, even the people of England would not be safe from its ravages*.

Once again, as in 1817-19, this indeed proved to be the case, as the *Liverpool Mercury* for 1847 has "**fever**" at 119 words/million with "*typhus*" as top collocate at log ratio 11.8. On the 18th June 1847, the *Liverpool Mercury*, while reporting on the sixth death of a Catholic priest from typhus in Liverpool within two months, opined that the cause was:

> *… the burden thrown by the national calamity on Liverpool …… the deluge of pauperism and disease which has been pored* [sic] *upon our shores from Ireland*.

The same was true in Scotland where the *Glasgow Herald* has "**fever**" at 52.2 words/million in 1847, with "*typhus*" as top collocate at log ratio 12.9 [see also 24]. However, on the 2nd July 1847 the *Glasgow Herald* was able to report that "*typhus fever is not at the present time alarmingly prevalent in London*", presumably accounting for the statistically significant signal for "**fever**" in 1847 for the newspapers of Belfast, Liverpool and Glasgow but not in *The Times*.

The following year the *Glasgow Herald* also scored another hit for "**fever**" at 39 words/million. "*Typhus*" was yet again the top collocate at log ratio 13.2 but followed closely by "*scarlet*" at log ratio 10.8. The succession of typhus by scarlet fever is illustrated by a report from 17th July 1848:

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 December 2022 doi:10.20944/preprints202012.0817.v2

> *Typhus fever, lately so prevalent, also steadily declines in frequency and fatality…. Scarlet fever is unusually prevalent among children, having almost entirely taken the place of measles and hooping-cough* [sic]*, which are comparatively rare at present. Smallpox still continues prevalent among the unvaccinated. Diseases of the brain and of the respiratory organs have been less fatal during June than during any previous month of the year*.

The latter comment on respiratory disease probably refers to the decline of the 1847-1848 influenza pandemic mentioned above. In 1849, "**fever**" at significant levels is confined to the *Belfast Newsletter* at 66.8 words/million. Among the collocates, we again find "*typhus*" at log ratio 14.1 but now also "*cholera*" at log ratio 8.0. Cholera is now known to be caused by *Vibrio cholerae*.

The next significant appearance of "**fever**" is in *The Era* for the years 1852 and 1853 at 23.1 and 21.4 words/million respectively. "*Typhus*" is again collocated at log ratio 13.1, along with "*yellow*" at log ratio 11.1 and "*cholera*" at log ratio 9.0. Yellow fever outbreaks in New Orleans [25] and the Caribbean are reported from November 1852 onwards, not decreasing until September 1853.

In 1870, "**fever**" appears at 68.1 words/million in the *Hampshire & Portsmouth Telegraph*. The collocates are various, in descending order of log ratio: "*enteric*" (13.1), "*scarlet*" (12.5), "*typhus*" (12.1), "*smallpox*" (11.4), "*typhoid*"(10.5), "*diphtheria*" (9.7), "*measles*" (8.6) and "*yellow*" (8.0). The yellow fever reports for 1870 were from Haiti, Brazil and Jamaica, different locations to those of the early 1850s. "*Enteric fever*" can be considered to be synonymous with typhoid – now known to be caused by *Salmonella typhi*. On 15th January 1870, while noting that scarlet fever was the leading infectious disease cause of death in a national survey of the previous week, closely followed by whooping cough and far ahead of measles, diphtheria and smallpox, the *Hampshire & Portsmouth Telegraph* reported that:

> *.. in the metropolis scarlet fever has been epidemic since 5th October last, but is now gradually subsiding.*

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 December 2022 doi:10.20944/preprints202012.0817.v2

However, by 15th October things were worse again:

> *The scarlet fever epidemic goes on increasing in fatality, the disease showing the most marked increase in London and in Bristol*.

In 1876, also in the *Hampshire & Portsmouth Telegraph*, "**fever**" reaches 59.3 words/million, with a very similar pattern of collocates but among which there are two new additions, "*scarlatina*" and "*gastric*". However, these are merely alternative names for scarlet fever and typhoid respectively. More importantly, in 1876 "*typhus*" is absent, unlike in 1870. Yellow fever again features in the foreign news, but in 1876 had switched its location once more, this time to Barbados, Brazil and Santo Domingo. A very similar pattern continued for the following two years in the *Hampshire & Portsmouth Telegraph*.

In 1881, *Reynold's Daily* significantly displays "**fever**" at 56.3 words/million with "*typhoid*", "*typhus*" and "*scarlet*" as collocates at log ratios 15.3, 11.8 and 10.2 respectively. An entry for 27th November 1881 describes how, in the London neighbourhood of:

> *… the parish of Marylebone, notorious for its fever dens … typhus fever of the most severe type was raging there, entirely attributable to the overcrowding of the people and bad sanitary state of the locality*

showing that in places, little had changed for much of the 19th century – a fact well recognised by campaigning authors of the time [e.g. 26].

In *The Times* for 1887, "**fever**" appears at 52.9 words/million, again with similar collocates to those found in the significant hits of the 1870s and earlier 1880s with the addition of "*malarial*" at log ratio 13.2. "*Scarlet*" is also collocated at log ratio 13.2, but in contrast to previous years, the discussion about scarlet fever is positive, for instance on the 21st January 1887 delivering the good news that:

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 December 2022 doi:10.20944/preprints202012.0817.v2

> *There has certainly been a steady and persistent decline of the London death-rate for this disease* [referring to scarlet fever] *for some years …* [and] *smallpox is for the moment almost stamped out in London, only five deaths having occurred during 1886*.

For the period of *"the Russian influenza"* pandemic, "**fever**" makes significant appearances in 1891 and 1892 in *The Times* at 51.8 words/million averaged over both years, and in the *Northern Echo* for 1891 at 122 words/million. The usual suspects are once again found as collocates, indicating that many other infectious diseases continued to make their way through the population in parallel to the ongoing pandemic of *"the Russian influenza"*. Interestingly, given the known ability of influenza to jump hosts from humans to pigs and vice versa, the *Northern Echo* reports on 16th April 1891 that:

> *swine fever is bad likewise, there having been 424 fresh outbreaks in Great Britain during the same period, with 2,119 pigs attacked … 917 diseased pigs were killed and 811 died in these five weeks only, 180 having recovered while 371 remained alive ailing*.

The *Ipswich Journal* continues to show a significant signal for "**fever**" for the years 1894 to 1896, with "*swine*" a collocate at log ratio 9.8. It was not until 26th December 1896 that the *Ipswich Journal* could report that "*swine fever is undoubtedly a diminishing quantity*". However in 1900 the same newspaper again scored "**fever**" at 59.6 words/million with "*swine*" a collocate at log ratio 9.8 and declaring on 26th May 1900 that "*the administrative county of West of Suffolk , is now an area infected with swine fever***".**

From 1900 to 1902, "**fever**" appears significantly in *The Times* averaging 96.7 words/million over the three years. Many of these entries are related to outbreaks of typhoid and blackwater fever – a complication of malaria [27] - among British troops in the Boer War.

Summary: search term "**fever**" is mostly used in the context of well-known diseases, including typhoid, typhus, scarlet fever and yellow fever or, in a veterinary context, "*swine fever*". Notably, when a simple reference is made to "*the fever*", without qualification, examination of the context often

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 December 2022 doi:10.20944/preprints202012.0817.v2

indicates typhus as the cause. For much of the early 19th century, for journalists writing in newspapers, "*the fever*" appears to be virtually synonymous with typhus. There is no evidence, using the present method, of any outbreak of feverish illness in the period from 1785 to 1890, which a contemporary newspaper felt it necessary to remark upon as unexplained.

Epidemic

The first year in which "**epidemic**" scores at statistically significant level is 1832 at 9.0 words/million in the *Liverpool Mercury*, referring overwhelmingly to cholera. This is followed by 1848 in which "**epidemic**" scores at 13.4 words/million in the *Glasgow Herald*. Although 1848 was an influenza pandemic year, there is no statistically significant collocation of "**epidemic**" with "*influenza*", although there is one reference to "*epidemic influenza*" on the 10th January 1848. "*Cholera*", by contrast, is collocated with "**epidemic**" at log ratio 9.25. On 4th December 1848 we find the *Glasgow Herald* indulging in some predictive epidemiology concerning the cholera outbreak, some optimism that things might improve when the weather got better, some observations on the greater burden borne by the poor and, finally, a look forward to what we would now call herd immunity:

> *Judging from the analogy of 1831 and 1832, the chief force of the epidemic will fall on the summer months of 1849. In the interim, judicious measures may destroy the pabulum of the disease; arrangements may be made usual for placing within the reach of the poor the resources of medical science; the seasons may interfere; and the half or two-thirds, of the population, who were before exposed to the disease, will not be likely to suffer a second time in the same proportion as before*.

1849, as the *Glasgow Herald* had predicted the previous year, indeed proved [28] to be another year when "**epidemic**" would reach significant proportions, at 14.0 words/million. On 30th March 1849, it reported the alarming statistic that:

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 December 2022 doi:10.20944/preprints202012.0817.v2

> *from the first outbreaking* [sic] *of the epidemic In Glasgow on the 14th day of Nov., till the 22nd of March, when it may be said to have almost ceased, there have been buried from cholera 3,777, or about 1.06 per cent of the whole estimated population*.

Further afield, the victims of the global cholera pandemic included the recently outgoing US President James K. Polk.

Similar statistical significance for "**epidemic**" is found in 1849 in *The Times* and *Liverpool Mercury*, at 16.0 and 9.27 words/million respectively. On 5th Sept 1849, *The Times* reported that:

> *The deaths registered in London in the week ending Sept. 1 were 2,796, of which 1,663 were by cholera, 234 by diarrhoea. The mortality exceeds that of any previous week. The greatest number ever registered before in any week since 1840 was 2,454 deaths, in the week ending Dec. 4, 1847, when the last epidemic of influenza prevailed. ….. The mortality is nearly three times the average of the season, and is sensibly felt all over the metropolis*.

1853 is the next year when "**epidemic**" reaches significant levels at 8.47 words/million. "*Cholera*" is a collocate at log ratio 9.65 in *The Times*, which reported on the 2nd December 1853 on the previous appointment of: "*Tuesday, the 29th of November, as a day of humiliation and prayer on account of the present visitation of epidemic cholera".* The *Hampshire & Portsmouth Telegraph* also scored "**epidemic**" at 9.98 words/million for that year with "*cholera*" collocated at log ratio 10.7. The following year, *Reynold's Daily* scored "**epidemic**" at 17.3 words/million. This was 1854, when the famous Broad Street pump episode took place [29].

In the next decade, 1865 appears with significant hits for "**epidemic**" in the *Glasgow Herald* and *Belfast Newsletter* at 10.4 and 13.5 words/million respectively. The *Belfast Newsletter* reported on 28th January 1865 that:

> *For some time past a fever epidemic has been raging in Liverpool, and Government recently sent down Dr. Buchanan* [Sir George Buchanan (1831-1895), later Chief Medical

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 December 2022 doi:10.20944/preprints202012.0817.v2

> Officer during "the Russian influenza" pandemic] *to report upon its cause and spread …. Dr. Buchanan concludes that the disease is mainly attributable to drink and to overcrowding, rather than to other sanitary defects*.

On 10th April 1865, the *Belfast Newsletter* was also reporting on an outbreak of an unidentified disease in Russia and Germany:

> *There was then, neither plague nor cholera in those countries. At St. Petersburg an epidemic of recurrent fever has prevailed since Sept. last, and has rather increased than diminished of late. Not much is known as regards its origin, although there can be no doubt that the severe cold, the dram-drinking which is now carried to an unheard-of extent, the vile quality of the spirituous compound which is sold to the people, the bad black bread, the putrid water which is used for drinking, and the overcrowding in the miserable lodgings of the poor have done much towards rendering the epidemic severe*.

By contrast with the *Belfast Newsletter*, the significant collocates of "**epidemic**" in the *Glasgow Herald* in 1865 include the usual "*typhus*" and "*cholera*" at log ratios 11.9 and 9.58 respectively. However, "*Russian*" is also a collocate at log ratio 6.78, referring to the same outbreak as the *Belfast Newsletter*. On the 7th April 1865, the *Glasgow Herald* reassured its readers that:

> *Dr. Murchison, of the London Fever Hospital, in a letter to the Times, says, - The public need be under little apprehension as to the importation of the Russian epidemic into England*.

The following day it reported that question on the "*Russian epidemic*" had be asked in the House of Lords. More intriguingly, given the known neurological complication of some coronaviruses, such as OC43, on the 22nd April 1865 the *Glasgow Herald* quoted from another parliamentary debate:

> *… your Lordship will have gathered that, neither as regards the fevers which are present in St. Petersburg, nor as regards the nervous disease which is occurring in North Germany,*

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 December 2022 doi:10.20944/preprints202012.0817.v2

> *are the circumstances such as have on former occasions led to the adoption of quarantine by this country; that, as regards the importability of the nervous disease, our danger in communicating with the Baltic ports (unless there were movements of masses of infected population) is apparently nothing, or next to nothing*.

The "*Russian epidemic*" with its "*fever*" and "*nervous disease*" may never have arrived in the UK, notwithstanding the reports from January in the *Belfast Newsletter*, mentioned above. By the following year of 1866, the *Pall Mall Gazette* has "**epidemic**" at 22.9 words/million, but "*Russian*" is now not collocated, whereas "*cholera*" is, at log ratio 9.80. *The Times* also scored "**epidemic**" at 22.9 words/million and on the 19th August 1866 wrote that "*The Lancet has commenced an investigation into the epidemic of cholera in the East-end of London*" and the collocation of "**epidemic**" with "*cholera*" in *The Times* for 1866 is log ratio 13.1.

In 1884 the *Ipswich Journal* records "**epidemic**" at a statistically significant 13.5 words/million, reporting on 29th March 1884 that "*There is an epidemic of sore throats in London, which is spoiling the talk in both Houses of Parliament*", listing four affected senior politicians. However, the main collocation is with "*cholera*" at log ratio 11.0 with most references to cases abroad. There is also mention during the year of smallpox in India as well as local measles and whooping cough outbreaks.

From 1890-1892, "**epidemic**" scores significantly in several newspapers, mostly associated with *"the Russian influenza"* pandemic. For instance, *The Times* has it at 29.2 words/million over those three years with "*influenza*" collocated at log ratio 11.4. Other epidemics ran contemporaneously, as shown by collocates with "*cholera*" at log ratio 10.1, "*typhoid*" at 8.9, "*typhus*" at 8.6, "*smallpox*" at 8.3, "*diphtheria*" at 7.2 and "*scarlet*" at 6.0.

In 1896, the *Ipswich Journal* again recorded "**epidemic**" at a statistically significant 12.3 words/million. The top collocate is "*measles*" at log ratio 10.9, which resulted in the closure of schools in Suffolk for several weeks from February to April 1896. Diphtheria is also mentioned on several occasions, but the virtual absence of references to influenza and cholera is noticeable.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 December 2022 doi:10.20944/preprints202012.0817.v2

The two remaining instances when "**epidemic**" rises above statistical significance are in *The Times* in 1918 and 2001, at 9.64 and 10.7 words per million respectively. 1918 is not just concerned with "Spanish flu" since typhus and cholera are also collocates of "**epidemic**" (log ratios 12.5, 11.3 and 10.5 respectively). It was more than 80 years before "**epidemic**" next achieved statistical significance in *The Times*. In 2001, "**epidemic**" has collocates of "*foot-and-mouth*", "*measles*", "*AIDS*" and "*flu*" (appearing instead of "*influenza*") at log ratios 12.9, 9.4, 7.5 and 6.6 respectively.

Summary: just as "**fever**" seems mostly to refer to typhus in the 19th century, "**epidemic**" is mostly associated with cholera. The only hint of any epidemic of unknown aetiology for the period from 1785 to 1890, is the "*Russian epidemic*" of 1865 and its potential arrival in the UK in the form of the unspecified fever which hit Liverpool early in that year, investigated by the government's Dr Buchanan. 1884's "*epidemic of sore throats*" seems scarcely worthy of consideration, as its most likely cause is a bacterial infection similar to, but milder than, that which causes scarlet fever.

Discussion

McKie [1] popularised Vijgen et al.'s proposal [6] that, on the basis of a Bayesian phylogenetic reconstruction of the date of the most recent common ancestor of OC43 coronavirus and its nearest animal coronavirus relative, OC43 is a candidate for the aetiological agent of the pandemic known to history as *"the Russian influenza"*, conventionally regarded, largely on the evidence of seroarchaeology with some support from molecular phylogenetics, as having been caused by influenza H3N8 [13].

The theory of McKie and Vijgen et al. also raises the question of the possibility that the previous "official" influenza pandemics of 1830-1833 and 1847-1848 might have been caused not by influenza but by a coronavirus. There is some tentative seroarchaeological and molecular phylogenetic evidence that suggests that these pandemics may have been caused by influenzas H1N1 and H1N8 respectively, although the argument is necessarily, given the distance of time, not as compelling as

 doi:10.20944/preprints202012.0817.v2

that for influenza H3N8 as the causative agent of *"the Russian influenza"* [13]. Certainly, at the time, although there was no recourse to any method of diagnosis other than symptomology, these two early 19th century pandemics were seen as a disease – influenza – different to all the other respiratory diseases affecting the population, among which tuberculosis and bacterial bronchitis would have been prominent. Doctors, especially those in crowded, polluted cities, would have encountered respiratory conditions, both infectious and non-infectious, on a daily basis. They appear to have been quite confident that the events of 1830-1833 and 1847-1848 were caused by something different to the usual daily respiratory cases, something that they called influenza, and which had a clinical description fundamentally congruent with what we describe as influenza today [e.g. 30].

Nevertheless, it is notable that, in Parsons' [31] review of influenza in England and Wales from 1847 to 1892, he observes that annual mortality from influenza decreased from 460 deaths per million (dpm) in the pandemic year of 1848 to less than 10 dpm per annum average for the period 1874 to 1889 (excepting 1875 and 1879 when it was nearer 20 dpm). For the third quarter of the 19th century, influenza appears to have ceased to be a major cause of mortality. Among modern seasonal flu outbreaks in the UK, the particularly bad winter of 2017-2018 produced 15,969 deaths attributed to influenza, or 253 dpm [32]. Such a level was not seen for influenza in the entire inter-pandemic period of 1849-1889, according to Parsons. There is a possibility that Parsons' figures for the pandemic years of 1890-1892 are underestimated – his cumulative influenza mortality of 1270 dpm for *"the Russian influenza"* (Figure 1 of his paper) would give a total death toll due to influenza of just under 45,000 individuals based on a late 19th century UK population of 35 million. This is well under half of the final total of excess deaths estimated by the Registrar-General [2]. If pandemic influenza deaths were under-certified in 1890-1892, there may have been a similar systematic underestimate of influenza on UK death certificates in the latter part of the inter-pandemic period in the 19th century.

However, it still seems unlikely that there could have been any other widespread outbreak of novel respiratory disease during the inter-pandemic years of 1847 to 1892, without notice being taken, even

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 December 2022 doi:10.20944/preprints202012.0817.v2

if the medical culture of the time had started to downplay influenza as a diagnosis. Evidence of any widespread outbreaks of respiratory, potentially coronavirus, infection that might have been confused with influenza, as McKie [1] and Vijgen et al. [6] suggest for *"the Russian influenza"*, appears to be absent.

COVID-19 presents frequently with a fever [15] and the assumption is made here that any other new pandemic coronavirus would produce the same symptom, at least in its first wave. However, we find that references to "**fever**" in the 19th century are overwhelmingly associated with diseases that were already well characterised then and still easily recognisable today. For instance, scarlet fever [33], typhus [34], typhoid [35], diphtheria [36], smallpox [37] and yellow fever [38] all present with other distinctive signs in addition to fever, that make retrospective re-diagnosis of their occurrence as instances of coronavirus infection, highly implausible. The remaining cause of "**fever**" in 19th century newspapers was yellow fever, but in the UK this was confined to arriving seamen and was not onwardly transmissible [39].

A constant cough was another of the hallmarks of COVID-19 infection in its first wave [15]. However, the commonness of this symptom makes it especially challenging to identify significantly elevated usage in newspapers, especially when patent medicine advertising pollutes the signal. We noted in passing that 1860 was a year when advertisements for patent cough medicines appeared to be particularly prevalent. However, since there is no other statistical signal for that year from our other search terms, it may simply indicate a bad spell of seasonal colds. References to whooping cough are clustered in the early 1790s, the 1870s and 1890. Whooping cough is a very characteristic form of cough [40] and, once again, the periodic resurgences of this disease are unlikely to be misidentifications of coronavirus outbreaks.

On the 25th June 1854, *The Era* remarked that:

> *… we have driven off the … worst foes by quarantine and other regulations; vaccination has defeated the small-pox* [sic]*; …. Typhus, however, still contends with us, and cholera*"

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 December 2022 doi:10.20944/preprints202012.0817.v2

This serves as an illustration of the way that newspaper references to epidemics in the 19th century are overwhelmingly associated with typhus, cholera and, from mid-century onwards, also with typhoid fever. Table 1 highlights the period from the early 1830s to the early 1870s when cholera had its largest impact on the UK. Cholera pandemics were prolonged affairs, occurring in 1826-1837, 1846-1860 and 1863-1875, with the disease first arriving in the UK in December 1831 [41]. During the second pandemic, 1853-1854 [42] was the worst year in the UK and 1866 [43] during the third pandemic. Table 1 shows that "**epidemic**" retrieves the years 1832, 1853, 1854 and 1866, providing a satisfying demonstration that our method can detect events which we can subsequently verify in the medical literature of the time. This builds our confidence that the corpus linguistic approach would detect a "missing" coronavirus pandemic, had one really occurred during the period examined.

The best that our present technique may offer towards the identification of an epidemic that does not fall into any well recognised class of disease is the mysterious "*Russian epidemic*" of 1865, which was the subject of some interest in the newspapers, although it may not have arrived in the UK at all, regardless of what it was that required the urgent dispatch of the eminent trouble-shooter Dr Buchanan to Liverpool in January of that year. Its description in the *New York Times* on 5th May 1865 is confusing, suggesting it was a mixture of ongoing typhoid and typhus outbreaks with perhaps even bubonic plague overlapping. It certainly was not respiratory in nature [44]. Therefore, if the "*Russian epidemic*" of 1865 was a single disease, and even if it was genuinely a medical novelty, it cannot provide us with a candidate coronavirus pandemic.

Conclusions

We set out, in the spirit of McKie [1] and Vijgen et al. [6], to find candidate disease outbreak events in the 19th century that may represent coronavirus pandemics from which our four milder seasonal coronaviruses are descended, just as our seasonal flu strains are descended from pandemic influenzas. Although conclusive clinical evidence of the causative agents of the 19th century outbreaks that are

 doi:10.20944/preprints202012.0817.v2

conventionally accepted as influenza pandemics, is unlikely to be found within the limitations of modern molecular technology, it should be remembered that retrieving influenza genomes from the victims of the "Spanish flu" would once also have seemed mere science fiction. In the meantime, we admit to our scepticism concerning the reassignment of influenza pandemics to coronavirus, partly because influenza was recognised, from early in the 19th century, as a specific constellation of symptoms that were not to be confused with the myriad other respiratory infections common in the cold, damp and polluted environments of Victorian Britain. In any case, Parsons' [31] review strongly suggests that the inter-pandemic period of 1849 to 1889 was progressively devoid of serious clinical impact from influenza. If Parsons' observation is real, rather than a mere artefact of medical culture in recording the disease, it means that there are very few, if any, suggestions of an outbreak of sufficient magnitude to serve as a plausible candidate for a previously unrecognised coronavirus pandemic of the 19th century.

Our analysis suggests that, for the Victorians, the main epidemic threats – and/or perhaps their main media disease obsessions - were the dramatic intrusions of cholera and typhus. In the earlier period, we see smallpox also circulating, dropping out to a certain extent – defeated, as *The Era* optimistically proclaimed in 1854, by vaccination - and being replaced with typhoid in the second half of the century. Measles, scarlet fever, diphtheria and whooping cough were almost omnipresent, attacking mainly children. We see little evidence of Victorian doctors hesitating over any infectious disease diagnosis, with most of those diagnoses being things we would still recognise today. We see little or no evidence of Victorian journalists, other than in 1865, speculating over new mysterious infectious diseases. This leaves little scope for postulation of unexplained respiratory outbreaks that would fit the bill as candidate coronavirus pandemics. The "Russian epidemic" of 1865 cannot be considered a contender, as it is a diagnostic mess and, in any case, lacks the necessary respiratory component.

Acknowledgements

We thank the following colleagues at Lancaster University: Andrew Hardie for his patient assistance with our learner driving of CQPweb, Tony McEnery for pre-publication sight of his papers on optical character recognition and Elena Semino for introducing us to the wide and fascinating field of corpus linguistics.

Supporting Information

Raw data used for Table 1 are available from Dryad at doi:10.5061/dryad.9kd51c5md

References

1. McKie R. Did a coronavirus cause the pandemic that killed Queen Victoria's heir? The Observer. 2020.
2. Honigsbaum M. The 'Russian' influenza in the UK: lessons learned, opportunities missed. Vaccine. 2011;29 Suppl 2:B11-5 https://www.zora.uzh.ch/id/eprint/57354/4/Russian_Influenza_lessons_revised.pdf. Epub 2011/07/27. doi: 10.1016/j.vaccine.2011.03.063. PubMed PMID: 21757097.
3. Arnold C. Pandemic 1918. The Story of the Deadliest Influenza in History: Michael O'Mara Books; 2018.
4. Taubenberger JK, Reid AH, Krafft AE, Bijwaard KE, Fanning TG. Initial genetic characterization of the 1918 "Spanish" influenza virus. Science. 1997;275(5307):1793-6. Epub 1997/03/21. doi: 10.1126/science.275.5307.1793. PubMed PMID: 9065404.
5. BBC. Norway royal bloodline 'British': http://news.bbc.co.uk/1/hi/world/europe/3746596.stm

2004.
6. Vijgen L, Keyaerts E, Moes E, Thoelen I, Wollants E, Lemey P, et al. Complete genomic sequence of human coronavirus OC43: molecular clock analysis suggests a relatively recent zoonotic coronavirus transmission event. J Virol. 2005;79(3):1595-604. Epub 2005/01/15. doi: 10.1128/JVI.79.3.1595-1604.2005. PubMed PMID: 15650185; PubMed Central PMCID: PMCPMC544107.
7. Jacomy H, Fragoso G, Almazan G, Mushynski WE, Talbot PJ. Human coronavirus OC43 infection induces chronic encephalitis leading to disabilities in BALB/C mice. Virology. 2006;349(2):335-46. Epub 2006/03/11. doi: 10.1016/j.virol.2006.01.049. PubMed PMID: 16527322; PubMed Central PMCID: PMCPMC7111850.
8. Riski H, Hovi T. Coronavirus infections of man associated with diseases other than the common cold. J Med Virol. 1980;6(3):259-65. Epub 1980/01/01. doi: 10.1002/jmv.1890060309. PubMed PMID: 6262459; PubMed Central PMCID: PMCPMC7166474.
9. Erkoreka A, Hernando-Perez J, Ayllon J. Coronavirus as the Possible Causative Agent of the 1889-1894 Pandemic. Infect Dis Rep. 2022;14(3):453-69. Epub 2022/06/24. doi: 10.3390/idr14030049. PubMed PMID: 35735759; PubMed Central PMCID: PMCPMC9222826.
10. Brussow H, Brussow L. Clinical evidence that the pandemic from 1889 to 1891 commonly called the Russian flu might have been an earlier coronavirus pandemic. Microb Biotechnol.

2021;14(5):1860-70. Epub 2021/07/14. doi: 10.1111/1751-7915.13889. PubMed PMID: 34254725; PubMed Central PMCID: PMCPMC8441924.
11. Berche P. The enigma of the 1889 Russian flu pandemic: A coronavirus? Presse Med. 2022;51(3):104111. Epub 2022/02/07. doi: 10.1016/j.lpm.2022.104111. PubMed PMID: 35124103; PubMed Central PMCID: PMCPMC8813723.
12. Forni D, Cagliani R, Pozzoli U, Mozzi A, Arrigoni F, De Gioia L, et al. Dating the Emergence of Human Endemic Coronaviruses. Viruses. 2022;14(5). Epub 2022/05/29. doi: 10.3390/v14051095. PubMed PMID: 35632836; PubMed Central PMCID: PMCPMC9148137.
13. Worobey M, Han GZ, Rambaut A. Genesis and pathogenesis of the 1918 pandemic H1N1 influenza A virus. Proc Natl Acad Sci U S A. 2014;111(22):8107-12. Epub 2014/04/30. doi: 10.1073/pnas.1324197111. PubMed PMID: 24778238; PubMed Central PMCID: PMCPMC4050607.
14. Ramassy L, Oumarou Hama H, Costedoat C, Signoli M, Verna E, La Scola B, et al. Paleoserology points to Coronavirus as possible causative pathogens of the 'Russian flu'. Microb Biotechnol. 2022;15(7):1943-5. Epub 2022/04/07. doi: 10.1111/1751-7915.14058. PubMed PMID: 35384322; PubMed Central PMCID: PMCPMC9111311.
15. Cevik M, Bamford CGG, Ho A. COVID-19 pandemic-a focused review for clinicians. Clin Microbiol Infect. 2020;26(7):842-7. Epub 2020/04/29. doi: 10.1016/j.cmi.2020.04.023. PubMed PMID: 32344166; PubMed Central PMCID: PMCPMC7182753.
16. Hardie A. CQPweb - combining power, flexibility and usability in a corpus analysis tool. International Journal of Corpus Linguistics. 2012;17(3):380–409.
17. McEnery T, Baker H, Dayrell C. Working at the interface of hydrology and corpus linguistics: using corpora to identify unrecorded droughts in nineteenth century Britain. In: Egbert J, Baker P, editors. Using Corpus Methods to Triangulate Linguistic Analysis. New York: Routledge; 2019.
18. Gabrielatos C, McEnery T, Diggle PJ, Baker P. The peaks and troughs of corpus-based contextual analysis. International Journal of Corpus Linguistics. 2012;17(2):151 - 75. doi: https://doi.org/10.1075/ijcl.17.2.01gab.
19. Bonferroni CE. Teoria statistica delle classi e calcolo delle probabilita. Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze. 1936;8:3-62.
20. Brown W. Notice of the Late Influenza in Edinburgh. Edinb Med Surg J. 1835;43(122):26-32. Epub 1835/01/01. PubMed PMID: 30330386; PubMed Central PMCID: PMCPMC5780928.
21. Uscher-Pines L, Pines J, Kellermann A, Gillen E, Mehrotra A. Emergency department visits for nonurgent conditions: systematic literature review. Am J Manag Care. 2013;19(1):47-59. Epub 2013/02/06. PubMed PMID: 23379744; PubMed Central PMCID: PMCPMC4156292.
22. Anon. Jungle Fever in Calcutta. Ind Med Gaz. 1872;7(7):165.
23. Kidd WL. A Concise Account of the Typhus Fever at Present Prevalent in Ireland, &c. Edinb Med Surg J. 1818;14(54):144-58. Epub 1818/04/01. PubMed PMID: 30329830; PubMed Central PMCID: PMCPMC5754719.
24. Orr RS. Historical and Statistical Sketch of the Progress of Epidemic Fever in Glasgow during the Year 1847. Edinb Med Surg J. 1848;69(175):363-78. Epub 1848/04/01. PubMed PMID: 30330644; PubMed Central PMCID: PMCPMC5790875.
25. Anon. Report of the Sanitary Commission of New Orleans on the Epidemic Yellow Fever of 1853. Br Foreign Med Chir Rev. 1856;18(36):285-303. Epub 1856/10/01. PubMed PMID: 30164919; PubMed Central PMCID: PMCPMC5199662.
26. Monjaras JE. The Part which Pauperism Plays in Cities. Its Extinction is the Best Prophylactic Means against Typhus. Public Health Pap Rep. 1895;21:280-4. Epub 1895/01/01. PubMed PMID: 19600655; PubMed Central PMCID: PMCPMC2329180.
27. Thin G. The Parasite of Malaria in the Tissues in a Fatal Case of Blackwater Fever. Br Med J. 1899;1(2005):1325-7. Epub 1899/06/03. doi: 10.1136/bmj.1.2005.1325. PubMed PMID: 20758514; PubMed Central PMCID: PMCPMC2462397.
28. Adams AM. Report upon Cholera as It Appeared in the 17th District of the City Parish of Glasgow, during the Months of November, December, January, February, and March, 1848-49. Edinb

Med Surg J. 1849;72(181):285-314. Epub 1849/10/01. PubMed PMID: 30330809; PubMed Central PMCID: PMCPMC5792854.
29. Snow J. Cholera and the Water Supply in the South Districts of London in 1854. J Public Health Sanit Rev. 1856;2(7):239-57. Epub 1856/10/01. PubMed PMID: 30378891; PubMed Central PMCID: PMCPMC6004154.
30. Hall CR. A Sketch of the Late Influenza as It Appeared in the Agricultural Districts of Cheshire. Prov Med Surg J. 1844;8(21):315-8. Epub 1844/08/21. doi: 10.1136/bmj.s1-8.21.315. PubMed PMID: 20793522; PubMed Central PMCID: PMCPMC2558472.
31. Parsons HF. On the Distribution of the Mortality from Influenza in England and Wales during Recent Years. Trans Epidemiol Soc Lond. 1894;13:114-26. Epub 1894/01/01. PubMed PMID: 29419216; PubMed Central PMCID: PMCPMC5542517.
32. Moss JWE, Davidson C, Mattock R, Gibbons I, Mealing S, Carroll S. Quantifying the direct secondary health care cost of seasonal influenza in England. BMC Public Health. 2020;20(1):1464. Epub 2020/10/01. doi: 10.1186/s12889-020-09553-0. PubMed PMID: 32993588; PubMed Central PMCID: PMCPMC7526100.
33. Freeman R. Observations on an Epidemic Scarlet Fever and Sore Throat; Communicated to Dr. Bradley. Med Phys J. 1803;9(48):157-61. Epub 1803/02/01. PubMed PMID: 30491431; PubMed Central PMCID: PMCPMC5672358.
34. Mossman G. The History of a Remarkable Case of Typhus Fever, Immediately Succeeded by Measles, Terminating Successfully. Ann Med (Edinb). 1797;2:298-306. Epub 1797/01/01. PubMed PMID: 30299878; PubMed Central PMCID: PMCPMC5112482.
35. Anon. On the Analogies and Differences between Typhus and Typhoid Fever. Br Foreign Med Rev. 1839;8(16):428-47. PubMed Central PMCID: PMCPMC5621313.
36. Radcliffe JN. Diphtheria. Edinb Med J. 1859;4(9):862-3. Epub 1859/03/01. PubMed PMID: 29648233; PubMed Central PMCID: PMCPMC5303499.
37. Hebb CH. Mr. Hebb's Case of Small-Pox. Med Phys J. 1801;5(28):536-7. Epub 1801/06/01. PubMed PMID: 30489835; PubMed Central PMCID: PMCPMC5598689.
38. Clark J. A Treatise on the Yellow Fever, as It Appeared in Dominica in 1793 and 1796. Ann Med (Edinb). 1797;2:155-83. Epub 1797/01/01. PubMed PMID: 30299872; PubMed Central PMCID: PMCPMC5112476.
39. Milroy G. On Yellow Fever in Relation to the Home Population. Trans Epidemiol Soc Lond. 1869;3(Pt 1):32-47. Epub 1869/01/01. PubMed PMID: 29418950; PubMed Central PMCID: PMCPMC5523597.
40. Alderson J. On the Pathology of Hooping Cough. Med Chir Trans. 1831;16(Pt 1):78-93. Epub 1831/01/01. doi: 10.1177/09595287310160p109. PubMed PMID: 20895571; PubMed Central PMCID: PMCPMC2116631.
41. Anon. Dr. Marshall on the Cholera in Glasgow Med Chir Rev. 1832;16(31):130-4.
42. Traill T, Alison WP. The Outbreak of Cholera at Arbroath, in Scotland, in October, 1853. Br Foreign Med Chir Rev. 1854;13(25):298-300. Epub 1854/01/01. PubMed PMID: 30164418; PubMed Central PMCID: PMCPMC5185406.
43. Radcliffe JN. Report on the Recent Epidemic of Cholera (1865-1866). Trans Epidemiol Soc Lond. 1869;3(Pt 1):232-45. Epub 1869/01/01. PubMed PMID: 29418952; PubMed Central PMCID: PMCPMC5523599.
44. Anon. The official account of the Russian epidemic. Boston Med Surg J. 1865;72:377-82. doi: DOI: 10.1056/NEJM186506080721904.