Kronenberg, Christoph

Article — Published Version

A New Measure of 19th Century US Suicides

Social Indicators Research

Provided in Cooperation with:
Springer Nature

Suggested Citation: Kronenberg, Christoph (2021) : A New Measure of 19th Century US Suicides, Social Indicators Research, ISSN 1573-0921, Springer Netherlands, Dordrecht, Vol. 157, Iss. 2, pp. 803-815, https://doi.org/10.1007/s11205-021-02674-y

Social Indicators Research (2021) 157:803–815
https://doi.org/10.1007/s11205-021-02674-y

ORIGINAL RESEARCH

A New Measure of 19th Century US Suicides

Christoph Kronenberg[1,2,3]

Accepted: 18 March 2021 / Published online: 27 March 2021

Abstract
Suicides hurt families and the US economy with an annual cost of $69 billion. However, little is known about what determined suicide rates in the past. This is likely due to the lack of consistent data prior to the 20th century. In this article, I propose using newspaper suicide mentions for the period 1840–1910 as a proxy measure for suicide and perform several validation exercises. I show that the stylized facts like suicides drop during wars holds for suicide mentions. I also validate the newspaper suicide mentions against sparse suicide mortality data and a novel valence measure. This new measure can be used to assess the relationship between suicides and numerous policy changes happening in the 19th century that previously could not be explored. It thus offers a new research avenue for quantitative historical analyses, which can inform current policy via novel historical insights.

Keywords Suicide · Newspapers · Proxy

1 Introduction

Suicides are tragic events for witnesses, affected family, friends and for society in general. In order to understand the determinants of suicide, including studying the (un)intended consequences of past policies, data on suicides is necessary. Yet, currently there is only one data source for US suicides prior to 1900—the mortality schedules. Mortality schedules were parts of the census that asked for people who died the year before the census and took place in 1850, 1860, 1870, 1880, and 1885. The 1885 mortality schedules were part of some state censuses. States and territories are usually only covered in these mortality schedules once they have come under U.S. sovereignty or jurisdiction. This explains why many current states are not included in the data.[1] Furthermore, not all mortality schedules are digitized.

[1] The census bureau provides an overview of the mortality schedules by state and year: https://www.census.gov/history/pdf/mortality.pdf (last accessed 20th of July 2020).

✉ Christoph Kronenberg
christoph.kronenberg@uni-due.de

[1] CINCH, University of Duisburg-Essen, Berliner Platz 6-8, 45127 Essen, Germany

[2] Leibniz Science Campus Ruhr, Essen, Germany

[3] RWI – Leibniz-Institute for Economic Research, Essen, Germany

It is thus unsurprising that little is known about what determined suicide rates in the past. I offer a proxy measure for the state-year suicide rate by US state based on a large newspaper archive run by the Library of Congress. The proxy is built analogues to disease prevalence measures as mentions per 100,000 pages. However, alternatives are explored and in cases where only deaths are available as comparators mentions themselves are also used.

This paper provides evidence that suicide mentions are a feasible proxy for the suicide rate providing the first insight into suicide (mention) trends prior to 1900. The validity of suicide mentions as a proxy is shown by comparison with the sparse available mortality data and by comparing patterns with the subjective well-being measure developed in Hills et al. (2019).

The proxy measure, developed here, allows more detailed analysis as well as analysis for areas and years for which previously no data was available at all. The analysis here is mostly conducted on an aggregate level. However, future research could exploit that the newspaper data has exact dates and towns in which the newspaper was based.

2 Previous Work

Using text and especially newspapers as data is a recent development. Yet, already an entire literature using text as data exists (Algaba et al., 2020; Baker et al., 2016; Currie et al., 2020; Gentzkow et al., 2011, 2014, 2015; Gutmann et al., 2018; Marquardt, 2020). A prominent example is the work by Baker et al. (2016). They search ten newspapers, which are available from 1900 to today for a list of keywords indicating policy uncertainty and then divide those by total number of articles in the same newspaper and month. Yet, most of this literature focuses on how an aspect of the newspaper market is influenced by or influences societal changes (Gentzkow et al., 2011, 2014, 2015). An exception is the work by Bencsik (2020). She shows that crime distresses the neighborhoods in which they occur, but only after the crimes have been reported in newspapers.

Some work creates and validates a measures suicide, suicidality or well-being. Hills et al. (2019) create a measure of national subjective well-being. They use sentiment analysis to capture the subjective mental well-being of book authors using Google Books to create measures of national subjective well-being for the US, the UK, Germany and Italy.

Vandoros et al. (2019) show that economic uncertainty, as measured by mentions of certain macroeconomic and political terms in newspapers, could lead to increased suicides in the short-run.

Arendt (2018) shows for fifteen newspapers from Austria (1819–1944) that there is covariation between newspaper reports of suicides and actual suicides. He further shows that newspaper reports of suicides predict future suicides, but suicides do not predict future newspaper reports of suicides. Arendt (2019) explores whether suicide reporting in two Austrian newspapers in the years 1855, 1865, 1875 & 1885 was responsible given the current state of knowledge and finds that the level of responsible report was low and did not improve during the observation period. Arendt (2020) analyzes five newspaper each from a different territory of the Austro-Hungarian Empire between 1871 and 1910. He again finds covariation between the number of suicide reports in newspapers and the number of suicides within all five newspapers.

Several other projects have used newspapers as a data source to quantify historical events for which no records exist. For example Monkkonen (2006) creates a database of

Table 1 Years covered by each award

Award year	Years covered by award
2016–2019	1690–1963
2010–2015	1836–1922
2009	1860–1922
2008	1880–1922
2007	1880–1910
2006	–
2005	1900–1910

In the year 2006, no award was granted

murders in New York city by using newspaper reports in the New York Tribune, the largest contributing newspaper in this study. Atalay et al. (2020) digitized job postings in several US newspapers (Boston Globe, the New York Times, and the Wall Street Journal) from 1950 to 2000 to describe the evolution of work tasks to a previously impossible degree.

3 Data Sources

3.1 Digital Newspaper Program

The National Digital Newspaper Program (NDNP) is a collaboration between the National Endowment for the Humanities (NEH) and the Library of Congress (LC). Their website "Chronicling America" provides digitized historical newspapers. NEH-funded institutions (awardees) conducted the digitization and are spread across all US states, except Massachusetts. Awardees digitized 100,000 pages a year since 2005.[2] The historical years from which newspaper pages are digitized vary by year, but are in the range 1690–1963.

Table 1 provides the historical years covered by each grant e.g. the 2018 grants covered 1690–1963. Table 2 provides an overview of which states received grants in which year (2005–2019) and who the awardee in that state is.

3.2 Mortality Data

I merge two mortality datasets. First, I discovered the summary of the 1850 and 1870 mortality schedules and digitized them. They cover 31 and 44 states respectively. From 1900 onwards cause specific state-year mortality data has been made available on the NBER website[3] by Miller (2008). Haines (2001) describes the underlying recording system and describes that death registration similar to modern standards was only implemented in the 1930s. Thus, in 1900 only 10 states had data of adequate quality to report. Thus, comparing post-1900 mortality data to newspaper suicide mentions is not ideal as the newspaper data is fading out (see Fig. 1) and the mortality data is starting to fade in.

[2] Except for 2006, when no awardee was funded.

[3] The data is available at https://data.nber.org/data/vital-statistics-deaths-historical/ (last accessed 12th of May 2020).

Table 2 State award years

State	Awardees	Year (20XX)
AL	U. of Alabama, Tuscaloosa	18
AK	Alaska SL Historical Collections	18,16
AZ	Arizona Dept. of Libraries, Archives, & Public Records	17,12,10,08
AR	Arkansas SA	19,17
CA	U. of California, Riverside	18,15,09,07,05
CO	History Colorado	18,16
CT	Connecticut SL	19,17,15,13
DE	U. of Delaware	19,17,15
FL	U. of Florida, Gainesville	19,17,15,13,05
GA	Digital Library of Georgia	19,17
HI	U. of Hawai‘i at Mānoa	12,10,08
ID	Idaho State Historical Society	17,15,13
IL	U. of Illinois, Urbana	18,16,13,11,09
IN	Indiana SL	17,15,13,11
IA	State Historical Society of Iowa	16,14,12
KS	Kansas State Historical Society	13,11,09
KY	U. of Kentucky, Lexington	11,09,07,05
LA	Louisiana State U	13,11,09
ME	Maine SL	18,16
MD	U. of Maryland, College Park	18,16,14,12
MI	Central Michigan U	18,16,14,12
MN	Minnesota Historical Society	19,17,15,11,09,07
MS	Mississippi Dept. of Archives & History	17,15,13
MO	The State Historical Society of Missouri	12,10,08
MT	Montana Historical Society	18,13,11,09
NE	U. of Nebraska-Lincoln Libraries	18,16,09,07
NV	U. of Nevada, Las Vegas	18,16,14
NJ	Rutgers U. Libraries, the New Jersey SA & SL	18,16
NM	U. of New Mexico	14,12,10
NY	The New York Public Library, Astor, Lenox & Tilden Fdn	09,07,05
NC	U. of North Carolina, Chapel Hill	18,16,14,12
ND	State Historical Society of North Dakota	17,15,13,11
OH	Ohio History Connection	16,12,10,08
OK	Oklahoma Historical Society	13,11,09
OR	U. of Oregon	13,11,09
PA	Penn State U. Libraries, U. Park	12,10,08
RI	Providence Public Library	19
SC	U. of South Carolina	13,11,09
SD	South Dakota Dept. of Education	18,16,14
TN	U. of Tennessee	14,12,10
TX	U. of North Texas	16,11,09,07
UT	U. of Utah, Marriott Libraries	09,07,05
VT	U. of Vermont	14,12,10
VA	Library of Virginia	19, 17,14,09,07,05

Table 2 (continued)

State	Awardees	Year (20XX)
WA	Washington SL	18,12,10,08
WV	West Virginia U. Libraries	19,17,15,13,11
WI	Wisconsin Historical Society	19,17,15
WY	U. of Wyoming Libraries	19

New Hampshire and Massachusetts never received an award are thus not shown for brevity. The following abbreviations have been used for brevity: *SL* state library, *U*. University, *SA* State Archive

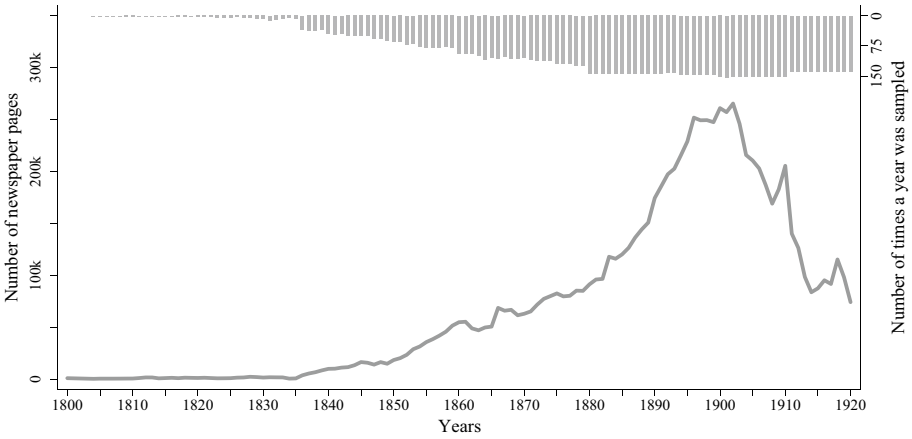

Fig. 1 Distribution of pages. *Note* the line shows the yearly number of observations per year. The spikes descending from the top of the graph show how often a historic year was sampled. For example, the drop in newspaper pages around 1899 is not due to the sampling of the NDNP, but due to a real reduction in newspaper pages

I lose observations in the comparison, because Massachusetts is available in both mortality datasets, but has never been part of the NDNP. On the other hand, states that joined the union only after 1920 are never available in the mortality data (e.g. Hawaii and Alaska joined the union in 1959).

3.3 Data Preparation

I first scrape the suicide mentions and the total number of pages available in the archive. I do not merge on the newspaper page level, due to a large share of missing page numbers, but first aggregate the pages and the suicide mentions to state-year level and then merge the suicide mentions to the total pages.

In the cleaning process, I dropped all observations that are not from current states but from current districts or territories—like Puerto Rico, Guam, etc. The reason they are dropped is, that NEH grants are only given to current-day states. Thus, it is unknown which state organization scanned those pages. The District of Columbia is also dropped, because the Library of Congress does the scanning, but does not report how many pages were scanned in which year. It is thus difficult to have any idea about the underlying data generating process.

I also only keep data from local newspapers in the sense that these newspapers were only distributed in the state they were published in. This should reduce over-counting due to celebrity suicides.

For the period 1800–1920 this leads to 932,457 suicide mentions and 8,677,032 newspaper pages. Aggregating this to state-year level there are 3,112 observations. The scraping is conducted in Stata 15.1 using the jsonio java plugin provided by Buchanan (2015). The aggregated state-year level data and the scraping code is available on my website at https://sites.google.com/view/christoph-kronenberg. Chronicling America also provides excellent resources on how to access the data at https://chroniclingamerica.loc.gov/about/api/.

4 Data Description

Table 3 answers the question how much of the newspaper market is represented in this sample. Unfortunately, that information is only available for 1840 and 1850 and thus comparison can only be made with newspapers from those years that are in the scraped dataset. Overall, the scraped dataset covers 3% of newspapers at the time. In a few cases the scraped dataset includes newspapers for states that were not part of the census newspapers count, see West Virginia (1850) and Hawaii (1840 and 1850).

In terms of contributions from individual newspapers, the top three are the New York Tribune with 130 thousand pages, the New York Herald with 116 thousand pages and the Evening star (Washington, D.C.) with 90 thousand pages. The mean/median number of pages per newspaper is 21/8 thousand for 1,684 newspapers.

Figure 1 displays the number of pages available in the data per year. The bars descending from the top indicate the number of times a historical year was covered by a NEH grant. This information is a combination of the information provided in Tables 1 and 2. For example, the combination Arizona and the year 1881 is coded as four. Newspapers from Arizona were scanned four times and each time 1881 was covered in the award period. 1879 on the other hand was coded as three, because the 2008 award only covers the period 1880–1922. The spikes follow the overall pattern of the pages and help to understand whether short-term fluctuations from the overall trend are due to sampling or real changes in the production of newspaper pages.

The number of pages drops around the turn of the century, despite the NEH grants being similar over that period. Thus, it is likely that this drop is a real change in newspaper production and not characteristics of the digitization process. A similar pattern can be observed for the American Civil War (1861–1865) for this period the number of pages' declines, while the number of times those years were sampled is increasing.

Figure 1 shows that newspaper pages prior to 1840 are scarce. This appears partly driven by the number of grants covering those years; see top part of the figure. However, Dill (1928) reports that in 1800 150 newspapers were active in the US increasing to 393 in 1810, 861 in 1820, 1300 in 1830 and 1403 in 1840. However, the number of copies took a little bit longer to increase as in 1810 22.5 million copies were produced increasing to 68.1 million in 1828, 90.4 million in 1835 and 195.8 million in 1840. Prior to these changes newspapers were so expensive that only a small number was produced and read. The current day price of a pre-1840 newspaper is estimated to be around $20.[4] The introduction of

[4] See the webpages of Encyclopedia Britannica and University Library—University of Illinois at Urbana-Champaign (both last accessed 25th of February 2019).

Table 3 Newspaper sample versus newspaper census

Newspapers in State	1840		1850	
	Total	In sample	Total	In sample
Alabama	28	0	60	0
Arkansas	9	0	9	0
California	0	0	7	0
Connecticut	44	0	44	0
Delaware	8	0	10	0
Florida	10	0	10	0
Georgia	40	0	51	0
Hawaii	0	1	0	1
Illinois	52	1	107	1
Indiana	76	2	107	3
Iowa	4	2	29	4
Kentucky	46	0	62	0
Louisiana	37	1	55	5
Maine	41	0	49	1
Maryland	49	1	68	2
Michigan	33	0	58	1
Mississippi	31	11	50	9
Missouri	35	2	61	3
New Jersey	40	0	51	0
New York	302	1	428	2
North Carolina	29	4	51	9
Ohio	143	5	261	12
Pennsylvania	229	5	309	8
Rhode Island	18	0	19	0
South Carolina	21	3	46	5
Tennessee	56	0	50	2
Texas	0	0	34	1
Vermont	32	9	35	9
Virginia	56	2	87	2
West Virginia	23	3	0	2
Wisconsin	6	0	46	1
New Mexico	0	0	2	0
Oregon	0	0	2	0
Sum	1498	50	2258	80
Percent		3.34%		3.54%

The 1840 and 1850 number of newspapers in the US is based on the Sixth and Seventh Census of the United States. The information is provided by the university library of University of Illinois at Urbana-Champaign and is available at https://guides.library.illinois.edu/antebellum-american-newspapers (Last accessed 17th of April 2019). New Hampshire and Massachusetts never received an award, see Table 3, and are thus not included in this table

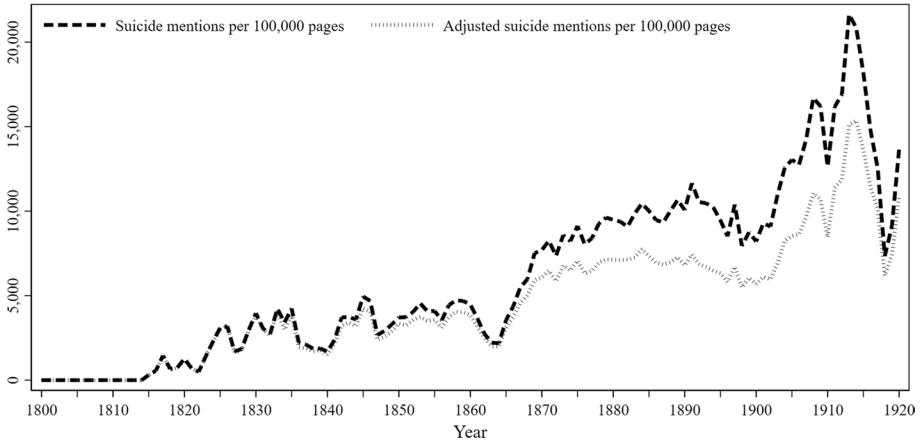

Fig. 2 Suicide mentions per 100,000 pages over time. *Note* the dashed line shows the yearly number of suicides mentions per 100,000 pages per year. The adjusted measure (dotted line) only counts at most one suicide mention per newspaper/day before aggregation. This should reduce the potential problem of over counting of celebrity suicides. The x-axis shows the years from 1800 to 1920, ticks indicate five-year intervals

new printing technologies combined with the use of steam to power printing presses and increased literacy led to so-called penny presses cheap tabloid newspapers.

I search this archive for occurrences of the term "suicide(s)". The result is a dataset of all mentions of suicides within the archive. Given that I have more pages for some states and years than for others, I divide the total number of suicide-mentions by state and year by the total number of pages per state-year cell and multiply by 100,000. This approach is analogues to cause-specific mortality that is reported as deaths per 100,000 individuals for a specific disease/cause of death (e.g. suicide).

Figure 2 shows the trend in suicide mentions per 100,000 pages over time. The civil war dip is far larger than in Fig. 1, which suggests that the trend in suicide mentions is far larger than purely explained by the trend in pages. A similar finding has been reported by Wasserman et al. (1994) looking at cover pages of the New York Times. Yet, they also show that the editor drove this pattern for the New York Times.

Figure 2 also shows the comparison between the raw suicide mentions per 100,000 pages and only counting one mention per day and newspaper. The latter should reduce concerns about over-counting suicides from well-known individuals. The lines are virtually identical until 1865 and then suicide mentions grows faster for a decade after which the two lines continue parallel of each other. This indicates that spikes in suicide mentions are not a huge concern, at least when looking at aggregate trends.

5 Data Validation

The natural validation process would be to compare the newspaper suicide mentions to the real suicide prevalence. Yet, as outlined in 3.2 this data is only available for 1850, 1870 and 1900- and even then lack states before they join the union. Data quality issues have also been raised (Haines, 2001), which is unsurprising given that they are large data quality issues even today (Fernandez, 2019).

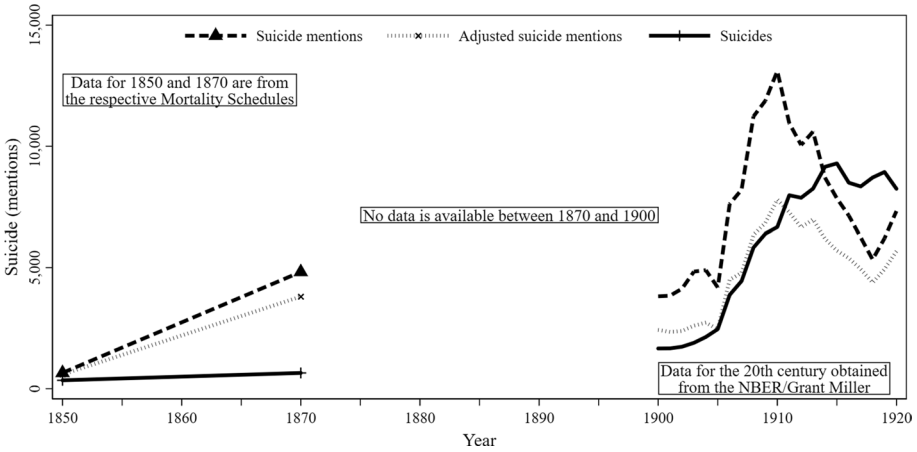

Fig. 3 Suicide mentions versus suicides 1850, 1870 and 1900–1920. *Note* The connected lines between 1850 and 1870 show suicides reported in historical mortality schedules for 22 states (1850) and 33 states (1870). The lines from 1900 onwards show suicides as provided by Grant Miller on the NBER website https://data.nber.org/data/vital-statistics-deaths-historical/ (last accessed 20th of July 2020)

I still compare the suicide mentions proxy against suicide prevalence data post 1900 in order to show that they despite all outlined problems compare reasonably well.

Figure 3 compares 1850, 1870 and 1900–1920 suicide counts to suicide mentions and the adjusted suicide mention count. Given the unavailability of year population numbers counts of suicides and mentions are compared. Furthermore, due to unavailable suicide counts for some states this comparison can only be conducted for 32 states from 1900 onwards.[5] For 1850, no suicide data is available from Hawaii, Minnesota and West Virginia as these states were only admitted to the union later, the same holds for Hawaii in 1870.

For 1850 and 1870 suicide mentions appear to over-count suicides. Yet it is also possible that the suicide count is an undercount given that the mortality schedules are based on the reporting of relatives. The adjusted suicide mentions track the actual number of suicides fairly well from 1900 to 1910 after which suicides grow while adjusted suicide mentions decline. The trends for the raw suicide mentions are similar at higher levels. Thus, it appears that newspaper suicide mentions, raw or adjusted, are either a reasonably good proxy or an undercount of the true number of suicides. Yet, the newspaper data includes data for another 15 states[6] post 1900 not covered in the data provided by Grant Miller and the NBER.

Figure 4 compares the suicide mentions per 100,000 pages versus the valence measure put forth in Hills et al. (2019). Their measure captures subjective wellbeing as implied by the sentiment (or valence) in serval million books. While subjective wellbeing or valence are not identical with suicide (mentions) there is a known strong

[5] These 32 states are Connecticut, Indiana, Maine, Michigan, New Jersey, New York, Vermont, California, Colorado, Maryland, Pennsylvania, Washington, Wisconsin, Ohio, Minnesota, Montana, North Carolina, Utah, Kentucky, Missouri, Virginia, Kansas, South Carolina, South Dakota, Tennessee, Illinois, Louisiana, Oregon, Delaware, Florida, Mississippi and Nebraska.

[6] These 15 states are Alabama, Alaska, Arizona, Arkansas, Georgia, Hawaii, Idaho, Iowa, Nevada, New Mexico, North Dakota, Oklahoma, Tennessee, Texas and West Virginia.

Fig. 4 Suicide mentions versus valence measure from Hills et al. (2019). *Note* The black dashed line shows the yearly valence as provided by Hills et al. (2019). The dashed line shows the number of suicide mentions per year and the dotted line the adjusted measure. The latter only counts at most one suicide mention per newspaper/day before aggregation. The vertical lines indicate the US Civil War and the First World War in line with Fig. 5 in Hills et al. (2019)

correlation between subjective measures of wellbeing and quality of life with suicides (Eckersley & Dear, 2002; Hays & Fayers, 2020; Helliwell, 2006). The valence measure drops during both the Civil War and the First World War as do the suicide mentions. The valence measure declines already prior to both wars, whereas the suicide mentions only declines around the beginning of both wars.

The solid line shows the yearly number of suicides. The dashed line shows the number of suicide mentions per year and the dotted line the adjusted measure. The latter only counts at most one suicide mention per newspaper/day before aggregation.

Another aspect of validation is whether suicide mentions are indeed, capturing suicides and what share of suicides are captured. For this suicide mentions from Maine 1850, 1870 and 1900–1920 where hand-validated. There are 2,495 automatically recognized mentions, hand-validation shows that this is an 98% accuracy rate. In 45 cases the text recognition reported a suicide mention where there was none. However, of the remaining mentions of the word suicide, some covered suicidality, suicides in other countries, suicide attempts that did not lead to death or instances of the word suicide appearing without a suicide having taken place in the USA. Ignoring these cases 70% of all suicide mentions relate to suicides. In most of the cases, 1,619 to be exact, the name of the individual was reported, which allows me to check for the number of mentions per suicide. There are 1,113 unique names and thus 1.45 suicide mentions per suicide.

Figure 5 shows the share of suicide mentions after accounting for text recognition failures and invalid suicide mentions over suicide reports from the Mortality Schedules and Grant Miller. For 1870, 40% of all reported suicides are captured. From 1900 onwards the share drops as expected.

Figure 3 has shown that suicides went up in that time, while Fig. 1 has shown that the number of pages went down.

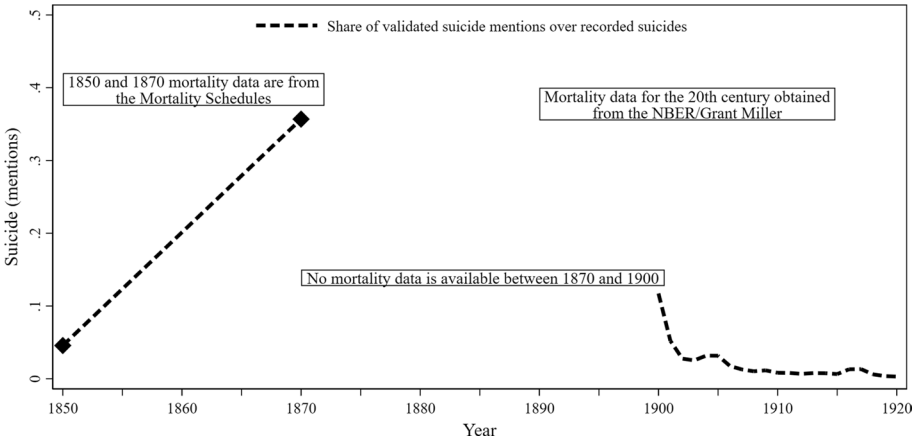

Fig. 5 valid suicide mentions as a share of reported suicides for maine. *Note* The 1850 and 1870 suicides are obtained from historical mortality schedules (1850) and (1870) as the denominator. The lines from 1900 onwards have suicides as provided by Grant Miller on the NBER website https://data.nber.org/data/vital-statistics-deaths-historical/ (last accessed 20th of July 2020) as the denominator

6 Discussion

I have extracted suicide mentions from nearly nine million newspaper pages from 1800 to 1920 and then validated them against the sparse existing data on suicide prevalence as well as a measure of subjective-wellbeing based on the sentiment in millions of books (Hills et al., 2019).

Thus, this work highlights the potential of using historical newspapers as a data source. There are of course challenges when working with data derived from newspapers published from 1800 to 1920. The scraping process is entirely machine-driven and thus some suicide mentions are a false positive. Yet, I have shown that 70% of all mentions relate to actual suicides in a subsample from Maine. False negatives (no reporting in newspapers despite suicides taking place) are a concern. However, the same is for any mortality data that is not based on coroner data. Thus, the same issue is present in the mortality schedules as relatives might misrepresent the cause of death of relatives for reasons of social prestige. Newspapers not reporting on suicides appears to have only taken place around the year 2000 [7]as can also be seen by the larger number of names reported in historical newspapers, which would be considered poor journalism currently.

Overall, it is unlikely that better data will ever be available for the 19th century and additional newspapers are continuously being added to the archive improving the quality. Yet using long historical datasets of any measure, including GDP, (un)employment or even population numbers should be done with caution. The same holds here, but having

[7] Today newspapers (often) carefully report about suicides or do not report them at all to prevent copycats. This reporting behavior likely arises from the 2001 joint-guidance by the Center for Disease Control and Prevention, the American Foundation for Suicide Prevention, the American Association of Suicidology and the Annenberg Public Policy Center on how to report suicides in media outlets. The WHO in 1999 offered guidance to media outlets in the SUPRE suicide prevention initiative. Prior to 1999, no guidance appears to have existed, especially not as far back as 1900. However, this blog post by Ritchie (2019) suggests that at least in the New York Times suicides are still over-represented by a factor of 7.

a dataset that has to be used with some caution is arguably better than having barely any data at all. However, I currently recommend using the data only from 1840 onwards given how rarely earlier years were sampled (see Fig. 1) and how few pages are available in the archive. I also currently recommend using the data only until 1910 given the sharp drop in available pages despite the reasonably high sampling rate.

This data resource offers researchers in the quantitative social sciences, linguistics and beyond the opportunity to study how numerous changes throughout the 19th century affected one of the currently leading causes of death (Hedegaard etal., 2018). There are many opportunities to study how changes in the economy, the westward expansion, slavery and the displacement of Native Americans, technological changes and wars affected suicides.

Future research could also attempt to extract data for other reported causes of death. Yet, it would need to be a cause of death that is often reported in newspapers and that all synonyms for the cause are known. While suicide as a word has been used for centuries without change, this is not true for most other causes of death.[8]

Acknowledgements I thank Daniel Avdic, Leah Boustan, Amitabh Chandra, Michael Darden, Katherine Eriksson, James Feigenbaum, Richard G. Frank, Vincent Geloso, Melissa McInerney, Martin Karlsson, Fabrizio Mazzonna, Christopher Ruhm, Colin Weiss and Nicolas Ziebarth for useful feedback. Conference/seminar participants at AUHE Leeds, CINCH, Essen Health Conference (2019), Essen Economics of Mental Health Workshop (2019), HEAL Lancaster, HESG 2019 (York), iHEA 2019 and School of Nursing and Health Sciences, University of Dundee provided useful comments. Cornelius Becker, Laura Bürger, Anne Günnel, Kai Miele and Felipe Oliveira provided valuable research assistance. All remaining errors are my own.

Funding Open Access funding enabled and organized by Projekt DEAL. I gratefully acknowledge funding from the Bundesministerium für Bildung und Forschung (Federal Ministry of Education and Research) Förderkennzeichen (Grant Number):01EH1602A via institutional funding for CINCH (Competent IN Competition and Health), Leibniz Science Campus Ruhr (http://www.lscr.de/) as well as conference funding from the German Health Economics Association (DGGÖ).

References

Algaba, A., Ardia, D., Bluteau, K., Borms, S., & Boudt, K. (2020). Econometrics meets sentiment: An overview of methodology and applications. *Journal of Economic Surveys*. https://doi.org/10.1111/joes.12370.

Arendt, F. (2018). Reporting on suicide between 1819 and 1944. *Crisis*. https://doi.org/10.1027/0227-5910/a000507.

Arendt, F. (2019). Assessing responsible reporting on suicide in the nineteenth century: Evidence for a high quantity of low-quality news. *Death Studies*. https://doi.org/10.1080/07481187.2019.1626952.

Arendt, F. (2020). The press and suicides in the 19th century: Investigating possible imitative effects in five territories of the Austro-Hungarian Empire. *OMEGA-Journal of Death and Dying, 81*(3), 424–435.

[8] Blindness used to be called ablepsy, stroke apoplexy, typhus was either called jail fever, ship fever or camp fever.

Atalay, E., Phongthiengtham, P., Sotelo, S., & Tannenbaum, D. (2020). The Evolution of Work in the United States. *American Economic Journal: Applied Economics, 12*(2), 1–34. https://doi.org/10.1257/app.20190070.

Baker, S., Bloom, N., & Davis, S. (2016). Measuring economic policy uncertainty. *The Quarterly Journal of Economics, 131*(4), 1593–1636. https://doi.org/10.1093/qje/qjw024.

Bencsik, P. (2020). *Stress on the sidewalk: The mental health costs of close proximity crime*.

Buchanan, B. (2015). *JSONIO: Stata module for I/O operations on JSON data*. Retrieved from https://ideas.repec.org/c/boc/bocode/s458087.html

Currie, J., Kleven, H., & Zwiers, E. (2020). Technology and big data are changing economics: Mining text to track methods. *AEA Papers and Proceedings, 110*, 42–48. https://doi.org/10.1257/pandp.20201058.

Dill, W. A. (1928). *Growth of newspapers in the United States*. University of Kansas.

Eckersley, R., & Dear, K. (2002). Cultural correlates of youth suicide. *Social Science and Medicine, 55*(11), 1891–1904. https://doi.org/10.1016/S0277-9536(01)00319-7.

Fernandez, J. M. (2019). The political economy of death: Do coroners perform as well as medical examiners in determining suicide? *SSRN*. https://doi.org/10.2139/ssrn.3360656.

Gentzkow, M., Petek, N., Shapiro, J., & Sinkinson, M. (2015). Do newspapers serve the state? Incumbent party influence on the US press, 1869–1928. *Journal of the European Economic Association, 13*(1), 29–61. https://doi.org/10.1111/jeea.12119.

Gentzkow, M., Shapiro, J., & Sinkinson, M. (2011). The effect of newspaper entry and exit on electoral politics. *American Economic Review, 101*(7), 2980–3018. https://doi.org/10.1257/aer.101.7.2980.

Gentzkow, M., Shapiro, J., & Sinkinson, M. (2014). Competition and ideological diversity: Historical evidence from US newspapers. *American Economic Review, 104*(10), 3073–3114. https://doi.org/10.1257/aer.104.10.3073.

Gutmann, M. P., Merchant, E. K., & Roberts, E. (2018). "Big data" in economic history. *The Journal of Economic History, 78*(1), 268.

Haines, M. R. (2001). The urban mortality transition in the United States, 1800–1940. *Annales de démographie historique, 1*, 33–64.

Hays, R. D., & Fayers, P. M. (2020). Overlap of depressive symptoms with health-related quality-of-life measures. *PharmacoEconomics*. https://doi.org/10.1007/s40273-020-00972-w.

Hedegaard, H., Curtin, S. C., & Warner, M. (2018). Suicide mortality in the United States, 1999–2017. *NCHS Data Brief No. 330.* Retrieved from https://stacks.cdc.gov/view/cdc/60894

Helliwell, J. F. (2006). Well-Being and social capital: Does suicide pose a puzzle? *Social Indicators Research, 81*(3), 455. https://doi.org/10.1007/s11205-006-0022-y.

Hills, T. T., Proto, E., Sgroi, D., & Seresinhe, C. I. (2019). Historical analysis of national subjective wellbeing using millions of digitized books. *Nature Human Behaviour, 3*(12), 1271–1275. https://doi.org/10.1038/s41562-019-0750-z.

Marquardt, K. (2020). *Identifying Physician Practice Style for Mental Health Conditions*.

Miller, G. (2008). Women's suffrage, political responsiveness, and child survival in American history. *The Quarterly Journal of Economics, 123*(3), 1287–1327. https://doi.org/10.1162/qjec.2008.123.3.1287.

Monkkonen, E. H. (2006). Homicide: Explaining America's exceptionalism. *American Historical Review, 111*, 76–94. https://doi.org/10.1086/ahr.111.1.76.

Ritchie, H. (2019). Does the news reflect what we die from? *Our World in Data.* Retrieved from https://ourworldindata.org/does-the-news-reflect-what-we-die-from

Vandoros, S., Avendano, M., & Kawachi, I. (2019). The association between economic uncertainty and suicide in the short-run. *Social Science and Medicine, 220*, 403–410.

Wasserman, I. M., Stack, S., & Reeves, J. L. (1994). Suicide and the media: The New York Times's presentation of front-page suicide stories between 1910 and 1920. *Journal of Communication, 44*(2), 64–83.