Prehospital Cardiac Arrest Should be Considered When Evaluating Coronavirus Disease 2019 Mortality in the United States

Nick Williams[1]

[1] National Library of Medicine, Lister Hill National Center for Biomedical Communications, Bethesda, Maryland, United States

Methods Inf Med 2023;62:100–109.

Address for correspondence Nick Williams, PhD, NIH Main Campus, NIHBC 38A - Lister Hill BG RM 9N912A, 8600 Rockville Pike, Bethesda, MD 20892, United States (e-mail: nick.williams@nih.gov).

Abstract

Background   Public health emergencies leave little time to develop novel surveillance efforts. Understanding which preexisting clinical datasets are fit for surveillance use is of high value. Coronavirus disease 2019 (COVID-19) offers a natural applied informatics experiment to understand the fitness of clinical datasets for use in disease surveillance.

Objectives   This study evaluates the agreement between legacy surveillance time series data and discovers their relative fitness for use in understanding the severity of the COVID-19 emergency. Here fitness for use means the statistical agreement between events across series.

Methods   Thirteen weekly clinical event series from before and during the COVID-19 era for the United States were collected and integrated into a (multi) time series event data model. The Centers for Disease Control and Prevention (CDC) COVID-19 attributable mortality, CDC's excess mortality model, national Emergency Medical Services (EMS) calls, and Medicare encounter level claims were the data sources considered in this study. Cases were indexed by week from January 2015 through June of 2021 and fit to Distributed Random Forest models. Models returned the variable importance when predicting the series of interest from the remaining time series.

Results   Model r2 statistics ranged from 0.78 to 0.99 for the share of the volumes predicted correctly. Prehospital data were of high value, and cardiac arrest (CA) prior to EMS arrival was on average the best predictor (tied with study week). COVID-19 Medicare claims volumes can predict COVID-19 death certificates (agreement), while viral respiratory Medicare claim volumes cannot predict Medicare COVID-19 claims (disagreement).

Conclusion   Prehospital EMS data should be considered when evaluating the severity of COVID-19 because prehospital CA known to EMS was the strongest predictor on average across indices.

Keywords
► Random Forest
► COVID-19
► public health
► statistical models
► syndromic surveillance

Introduction

Creating long-term, multisource, national surveillance data services for emerging disease response is a complex topic to which coronavirus disease 2019 (COVID-19) has given new importance.[1–5] Public health emergencies responses seldom leave surplus time or resources to stand up novel methods and respond, further essentializing (specific) disease

preparedness.[6–8] More often than not epidemic response is managed using preexisting data services, often legacy data series from yesteryear's epidemics.[9–11] Epidemic preparedness in the United States is generally weak, and the COVID-19 response is largely drawn from preexisting pan-flu emergency plans.[12,13]

During a public health emergency, the clinical knowledge needed to respond is developed by case surveillance drawn from preexisting data series. COVID-19 has presented an unusual opportunity to evaluate agreement across surveillance efforts within the United States. The ability to detect clinical findings from surveillance nets and epidemiology methods which were not necessarily designed to detect them in meaningful ways is a high priority for the future management of emerging infectious diseases. Strikingly, the difference in COVID-19 mortality for severe acute respiratory syndrome (SARS)-impacted countries (China, South Korea, and Australia) versus the United States comes down to what emergency response plan was last implemented (SARS vs. swine flu) and the fitness of surveillance (case specific vs general population) rather than deeper cultural, economic, or racial differences, as have been proposed in popular media.[14–20]

Objectives

In this study, public health surveillance data are processed using a machine learning approach to discover the relative agreement of a surveillance event series when predicting surveillance event series. Toward objectives, this study seeks to assess the agreement between event series and contrast the value of traditional surveillance methods (death certificates, influenza, and respiratory infection claims volumes) with nontraditional sources such as national Emergency Medical Services (EMS) call volume data in the COVID-19 era in the United States.

Methods

Statistic of Interest

Variable importance is the statistic of interest in this study. Variable importance means that when predicting the dependent variable, an independent variable which is of comparatively higher predictive value (association) than another is of higher (predictive) use value. When considering high variable importance with weekly event series data, series which help the machine learning models learn, predict, or guess the correct dependent weekly event series could be cooccurring or mutually observed events. The high variable importance scores from different sources suggest that series observe the same real-world event across surveillance efforts as they support prediction better than noise and other candidate series (other independent variables).

Of special interest are "high variable importance and independent variables" from a different data source than the dependent variable. High same-source variables are most likely high in value because they are similarly distributed across study weeks to their parent–sister series and in turn are not necessarily interesting. A series of events can be said to

have "agreement value" if it has high statistical agreement with other series from a different source. Low statistical agreement suggests "out of era" events or events which are not driven by the same causes as other series considered here.

Toward noise and disagreement, influenza and respiratory infection claims volumes are considered below with COVID-19 claims volumes. Claims volumes are traditionally used in influenza surveillance. As a test of the efficacy of the models described here, COVID-19 volumes should be able to "outperform" influenza volumes as the COVID-19 era is largely understood to be influenza sparse. In this way respiratory and influenza events could be understood as a control arm as well as a model output of independent interest.

Data Sources

Medicare

Medicare provided three event series to this study. Medicare encounter-level claims through July, 2021 were sourced through the Chronic Conditions Warehouse (CCW). Records from 2015 through July 2021 were considered. Claims that contained influenza, COVID-19, or respiratory infection diagnostic code were enrolled. A series was generated for counts of distinct individuals within a series by calendar week. The Medicare-sourced series do not describe the duration of illness but the frequency of billing over time for distinct individuals. Medicare claims provided three series to this study, specifically "Influenza Diagnostic (DX) Codes," "COVID-19 DX Codes," and "(Viral) Respiratory Infection DX Codes" series. The viral respiratory series includes fever, bronchitis, viral lung infection, acute respiratory distress syndrome (ARDS), and pneumonia ICD10-CM codes. Procedure, HCPS, and CPT-4 codes were not considered.

The Centers for Disease Control and Prevention

The Centers for Disease Control and Prevention (CDC) provided five series for this study. COVID Deaths: COVID deaths are described as weekly data set which disambiguates the primary cause of death (COD) on Multiple Cause of Death Certificates (MCDC) received by the CDC within the given week. The dataset further describes secondary causes of death when COVID-19 diagnostic codes are present. The COD All Cause, COD COVID Primary, and COD COVID Secondary series in this study were learned from this data set. COVID deaths data were retrieved from: "https://data.cdc.gov/NCHS/Provisional-COVID-19-Deaths-by-HHS-Region-Race-and/tpcp-uiv5."

Excess Mortality: CDC evaluates "excess mortality' or death certificates above expectation where expectation means the three smallest death rates per state within a condition and calendar week.[21–27] These deaths are technically preventable because they are being prevented in real time in other states. The interpretation of excess mortality is a complex topic, and individuals who die in excess are not necessarily dying significantly before they would have died baring excess. Two study series are learned from this data set, Observed Deaths and Excess Deaths. Excessive deaths are produced using Farrington flexible methods.[28,29] Excess mortality data were retrieved from "https://data.cdc.

gov/NCHS/Excess-Deaths-Associated-with-COVID-19/xkkf-xrst" and "https://github.com/Mortality-Surv-and-Reporting-Proj/county-level-estimates-of-excess-mortality."

The National Emergency Medical Services Information System

The National Emergency Medical Services Information System (NEMSIS) provided five event series to this study. NEMSIS is a complex data center which collects data from state-level supervising EMS authorities.[30,31] NEMSIS is designed to support EMS outcomes research and complex, evidence-based-medicine research.[32] NEMSIS has a stable data model of EMS episode values which are collected for every emergency (911) call which is routed to an EMS in the United States. A weekly extract was created using NEMSIS OLAP cubes for 2014 to 2016 and 2017 present. The cardiac arrest (CA) subset which codes calls for arrests before and after EMS arrived on the scene was also extracted. "NEMSIS Calls," "NEMSIS Calls CA Yes," "NEMSIS Calls CA No," and "NEMSIS CA Prior" to arrival and "NEMSIS CA After" arrival of the EMS crew were learned from NEMSIS. NEMSIS data was retrieved from: "https://nemsis.org/view-reports/public-reports/ems-data-cube/."

Statistical Models

The 13 series sets were integrated into a single "cases per week" data model and processed using machine learning methods in h2o.ai (https://www.h2o.ai). Specifically, models were generated to learn the dependent to independent variable relationships across the series such that each series

weekly value was attempted to be learned (predicted) from all other weekly event series values. Each series took a turn being the dependent variable in a Distributed Random Forest (DRF) model.[33] R squares (r^2) for models as well as scaled variable importance in decision-making are described below in detail. Models were cross-validated five times each. Note each series was itself a model (being predicted) from other series for a total of 14 models (13 event series and the study week itself). The statistic of interest is the variable importance of an independent variable when attempting to predict the dependent variable within a DRF model.

Models considered any volume between January 1st, 2018 and July 1st, 2021. Raw case count values were used, neither log/lag modeling nor relative rates were considered. Note DRF transforms numeric values to a continuous distribution in preprocessing (before processing). The fitness of "week" of event most likely obscures or confounds episode attribution of count data model events as a case could be transported by EMS, bill Medicare and populate a CDC death certificate within a calendar week or over several months in the case of advanced life support. The models should not be used to model the epidemic but rather to assess the agreement within the implicit (pseudo-harmonized) time scales of the series.

Results

►Table 1 describes the event series, its data source, the specific data set name, the series extracted for this study, the time range, and the total events within the series of

Table 1 Series ranges and data sources

Source	Data set	Series	Start	Stop	Case weeks
Medicare	Patient Level Claims	Influenza Events	01-01-2015	6/31/2021	38,37,068
Medicare	Patient Level Claims	COVID Events	01-01-2015	6/31/2021	1,78,49,177
Medicare	Patient Level Claims	Respiratory Infection Events	01-01-2015	6/31/2021	14,07,77,208
CDC	Excess Deaths Associated with COVID-19	Total Weekly Deaths	01-01-2017	04-12-2021	1,50,66,215
CDC	Excess Deaths Associated with COVID-19	Weekly Excess Deaths	01-01-2017	04-12-2021	9,51,680
CDC	Provisional COVID-19 Deaths by HHS Region, Race, and Age	Weekly MCDC	01-01-2015	11-12-2021	1,95,69,921
CDC	Provisional COVID-19 Deaths by HHS Region, Race, and Age	Weekly COVID Primary MCDC	01-01-2015	11-12-2021	5,90,090
CDC	Provisional COVID-19 Deaths by HHS Region, Race, and Age	Weekly COVID Secondary MCDC	01-01-2015	11-12-2021	6,52,472
NEMSIS	OLAP Cube	EMS Calls	01-01-2014	10-12−2021	23,79,08,326
NEMSIS	OLAP Cube	EMS Cardiac Arrest Calls	01−01−2014	10−12−2021	21,78,494
NEMSIS	OLAP Cube	EMS Non-Cardiac Arrest Calls	01−01−2014	10−12−2021	16,36,24,383
NEMSIS	OLAP Cube	All Cardiac Arrest Pre-EMS Arrival	01−01−2014	10-12-2021	19,10,767
NEMSIS	OLAP Cube	All Cardiac Arrest Post-EMS Arrival	01-01-2014	10-12-2021	2,67,727

Abbreviations: CDC, The Centers for Disease Control and Prevention; COVID-19, coronavirus disease 2019; EMS Emergency Medical Services; MCDC, Multiple Cause of Death Certificates; NEMSIS, The National Emergency Medical Services Information System.
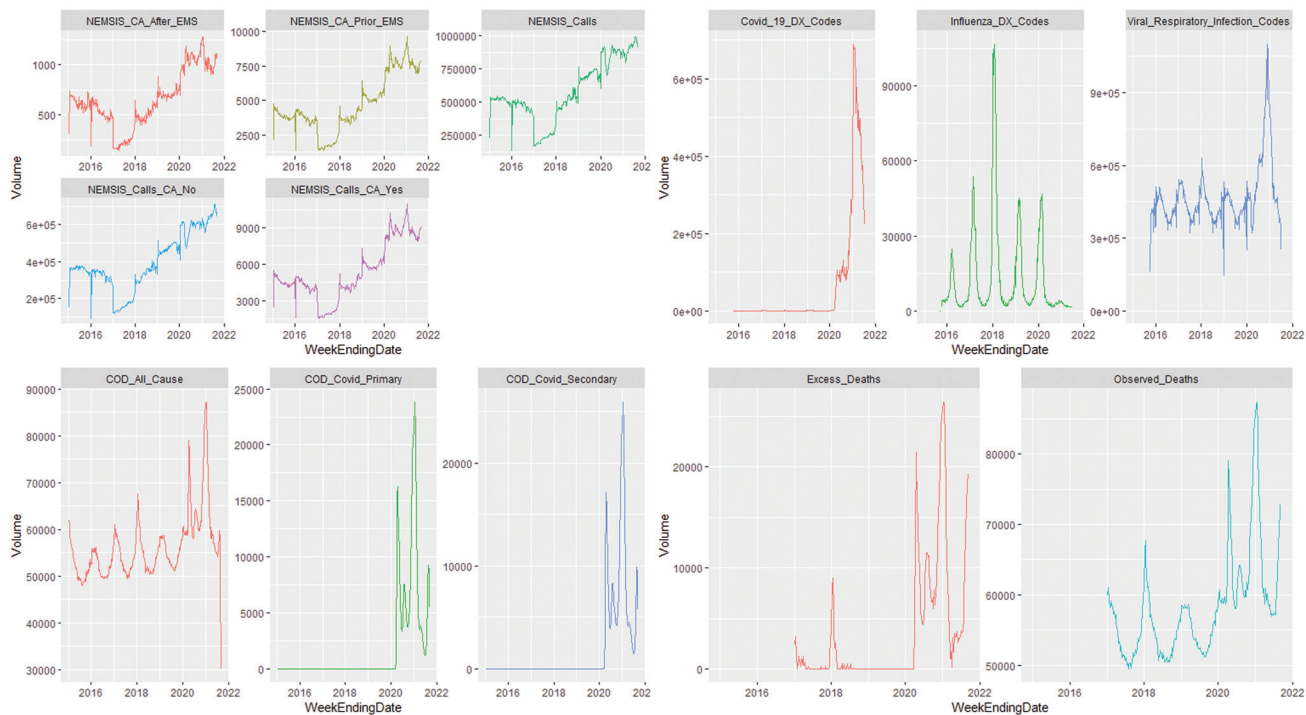
Fig. 1 The weekly event volume by event type. The upper right line graphs describe the per member per weekly occurrence of qualifying diagnostic codes on identifiable Medicare claims. COVID-19 (*red*), influenza (*green*), and respiratory infection codes (*blue*) are featured. The bottom right figures show the Excess Deaths (*Red*) and Observed Deaths (*Blue*) from which excess deaths are learned in the CDC excess mortality model. The upper left region describes the NEMSIS series with cardiac arrest after EMS arrival (Red), cardiac arrest prior (Brown), total calls (Green), calls without cardiac arrest (*Blue*) and calls with arrests (*Purple*). The lower left shows the all-cause mortality multiple cause of death certificate volumes (Red) and volumes where the primary (*Green*) and secondary causes of death (*Green*) were COVID-19. The *x*-axis is the study week, and the *y*-axis is the volume for all figures.

interest. Note that NEMSIS CA status is a declaration aggregate, and call where CA did not occur is a call with an explicit declaration. In turn, the total calls (sum) do not reflect the sum of CA and non-CA calls.

►**Fig. 1** shows the weekly volume of events within series described as totals in Table 1. The upper right describes Medicare weekly case events, and the bottom right describes excess mortality series. The upper left describes NEMSIS series, and the bottom left describes COVID-19 death certificates. Figure one demonstrates a collapse in influenza Medicare claims and spikes in covid and viral respiratory infection codes toward the end (right) of the series. COVID excess deaths and MCDC indicate similar peaks on the right side of the *x*-axis as well. All NEMSIS call volumes are elevated as time progresses.

►**Table 2** presents a matrix of dependent and independent variable series relationships, where the scaled variable importance is presented. Each column is a DRF model where the column header is the dependent variable. The independent variables are listed along the left-hand side of the table. In scaled variable importance measures, "1" is the highest value and independent variable can receive; and only one "1" can be awarded within a model. For example, dependent "Influenza DX Codes" weekly values from Medicare were most strongly learned from "Respiratory Codes" (1) from Medicare followed by "All Cause COD" (0.7191) from MCDC, "Observed Deaths" from Excess Deaths (0.6552) and "COVID-19 DX

Codes" from Medicare (0.4475). Alternately, "COVID 19 DX Codes" from Medicare shows "Week Ending Date" (1), followed by "COVID Primary COD" (0.4015) and "COVID Secondary COD" from MCDC (0.3455), "Excess Deaths" (0.2451), and strikingly "NEMSIS CA Prior EMS" (0.2445). Note that when predicting "COVID 19 DX Codes," "Respiratory Codes" are of little help (0.0636) but when predicting "Respiratory Codes," "COVID 19 DX Codes" are fairly helpful (0.8722) when making said prediction. r^2 is plotted above the dependent variable.

►**Table 3** replots ►**Table 2** values as above or below the model run's geometric mean variable importance score (column-wise geometric mean). The regions within the black outlines should be understood as variables from the same series source. While the models did know weekly features from the same data source their importance toward the study objective is minimal. For example, the only "same source series" variable importance below average was the Medicare "COVID 19 DX" model with influenza and viral respiratory variables being low importance (as expected). This should mean that the model did not learn what the weekly "COVID 19 DX Codes" volume was from viral infection and influenza codes; their series are independent in this study. Above variable importance within column models from different series should detail the interrelatedness of the multiseries weekly events. For example, "NEMSIS CA After EMS" shows above the geometric mean of variable

Table 2 Variable importance matrix and original values with dependent variables (column wise)

r²:	0.5532	0.9951	0.9968	0.9963	0.9839	0.9961	0.8897	0.7858	0.9048	0.9844	0.9823	0.8768	0.9537	0.9758
	Week Ending Date	NEMSIS Calls	NEMSIS Calls CA Yes	NEMSIS Calls CA No	NEMSIS CA After EMS	NEMSIS CA Prior EMS	Influenza DX Codes	Respiratory Codes	COVID 19 DX Codes	COD COVID Primary	COD COVID Secondary	COD All Cause	Excess Deaths	Observed Deaths
Week Ending Date	NA	0.1163	0.0904	0.0969	0.1025	0.1128	0.1705	0.7063	1	0.066	0.0703	1	0.1654	1
NEMSIS Calls	0.0115	NA	0.3877	0.8451	0.3234	0.3354	0.1749	0.0374	0.1052	0.0083	0.0149	0.0037	0.0097	0.045
NEMSIS Calls CA Yes	0.0141	0.908	NA	0.2401	0.6746	0.6455	0.099	0.0239	0.0517	0.1791	0.1865	0.0373	0.1637	0.0037
NEMSIS Calls CA No	0.0147	0.5712	0.1672	NA	0.3221	0.3379	0.2451	0.048	0.138	0.0249	0.0229	0.0054	0.0108	0.0143
NEMSIS CA After EMS	0.0039	0.1973	0.3163	0.1875	NA	1	0.1404	0.0282	0.0356	0.5651	0.585	0.0173	0.7209	0.0016
NEMSIS CA Prior EMS	0.0037	1	1	1	1	NA	0.1109	0.0482	0.2445	0.1867	0.1851	0.0236	0.3016	0.0041
Influenza DX Codes	0.1255	0.0056	0.0009	0.006	0.0035	0.0013	NA	1	0.0408	0.0095	0.0088	0.0789	0.0215	0.0025
Respiratory Codes	0.0834	0.0061	0.0009	0.006	0.0041	0.0009	1	NA	0.0636	0.0066	0.0108	0.1659	0.0575	0.003
COVID 19 DX Codes	0.0618	0.0512	0.0503	0.0444	0.0599	0.0599	0.4475	0.8722	NA	0.0461	0.0431	0.0702	0.0396	0.1613
COD COVID Primary	1	0.0676	0.0647	0.0615	0.0822	0.0774	0.2703	0.1078	0.4015	NA	1	0.4791	0.7849	0.0373
COD COVID Secondary	0.4674	0.0029	0.002	0.0031	0.0069	0.0063	0.3611	0.0567	0.3455	1	NA	0.6306	1	0.0704
COD All Cause	0.8665	0.0056	0.0049	0.0054	0.0088	0.0078	0.7191	0.2043	0.0726	0.8997	0.9163	NA	0.1388	0.1097
Excess Deaths	0.0195	0.0293	0.029	0.0286	0.0409	0.035	0.4472	0.3178	0.2451	0.8894	0.8964	0.0759	NA	0.4037
Observed Deaths	0.0073	0.0105	0.0078	0.0057	0.0162	0.0126	0.6552	0.0834	0.0516	0.666	0.6915	0.0871	0.9865	NA

Abbreviations: COD, cause of death; COVID-19, coronavirus disease 2019; EMS Emergency Medical Services; NEMSIS, The National Emergency Medical Services Information System.

Table 3 Variable importance matrix by dependent value column wise with independent variables above and below the geometric model mean (column wise)

	Week Ending Date	NEMSIS Calls	NEMSIS Calls CA Yes	NEMSIS Calls CA No	NEMSIS CA After EMS	NEMSIS CA Prior EMS	Influenza DX Codes	Respiratory Codes	Covid 19 DX Codes	COD COVID Primary	COD COVID Secondary	COD All Cause	Excess Deaths	Observed Deaths
Week Ending Date	NA	ABOVE	ABOVE	ABOVE	ABOVE	ABOVE	BELOW	ABOVE	ABOVE	BELOW	BELOW	ABOVE	ABOVE	ABOVE
NEMSIS Calls	BELOW	NA	ABOVE	ABOVE	ABOVE	ABOVE	BELOW	BELOW	BELOW	BELOW	BELOW	BELOW	BELOW	ABOVE
NEMSIS Calls CA Yes	BELOW	ABOVE	NA	ABOVE	ABOVE	ABOVE	BELOW	BELOW	BELOW	ABOVE	ABOVE	BELOW	ABOVE	BELOW
NEMSIS Calls CA No	BELOW	ABOVE	ABOVE	NA	ABOVE	ABOVE	BELOW	BELOW	ABOVE	BELOW	BELOW	BELOW	BELOW	BELOW
NEMSIS CA After EMS	BELOW	ABOVE	ABOVE	ABOVE	NA	ABOVE	BELOW	BELOW	BELOW	ABOVE	ABOVE	BELOW	ABOVE	BELOW
NEMSIS CA Prior EMS	BELOW	ABOVE	ABOVE	ABOVE	ABOVE	NA	BELOW	BELOW	ABOVE	ABOVE	ABOVE	BELOW	ABOVE	BELOW
Influenza DX Codes	ABOVE	BELOW	BELOW	BELOW	BELOW	BELOW	NA	ABOVE	BELOW	BELOW	BELOW	ABOVE	BELOW	BELOW
Respiratory Codes	ABOVE	BELOW	BELOW	BELOW	BELOW	BELOW	ABOVE	NA	BELOW	BELOW	BELOW	ABOVE	BELOW	BELOW
COVID 19 DX Codes	ABOVE	ABOVE	ABOVE	ABOVE	ABOVE	ABOVE	ABOVE	ABOVE	NA	BELOW	BELOW	ABOVE	BELOW	ABOVE
COD COVID Primary	ABOVE	ABOVE	ABOVE	ABOVE	ABOVE	ABOVE	BELOW	BELOW	ABOVE	NA	ABOVE	ABOVE	ABOVE	ABOVE
COD COVID Secondary	ABOVE	BELOW	BELOW	BELOW	BELOW	BELOW	ABOVE	BELOW	ABOVE	ABOVE	NA	ABOVE	ABOVE	ABOVE
COD All Cause	ABOVE	BELOW	BELOW	BELOW	BELOW	BELOW	ABOVE	ABOVE	BELOW	ABOVE	ABOVE	NA	ABOVE	ABOVE
Excess Deaths	BELOW	BELOW	BELOW	BELOW	BELOW	BELOW	ABOVE	ABOVE	ABOVE	ABOVE	ABOVE	ABOVE	NA	ABOVE
Observed Deaths	BELOW	BELOW	BELOW	BELOW	BELOW	BELOW	ABOVE	BELOW	BELOW	ABOVE	ABOVE	ABOVE	ABOVE	NA
Total Above	6	7	7	7	7	7	6	5	6	7	7	8	8	7

Abbreviations: COD, cause of death; COVID-19, coronavirus disease 2019; EMS Emergency Medical Services; NEMSIS, The National Emergency Medical Services Information System.

Table 4 Scaled variable importance above the geometric mean row wise (independent variable) across models (column wise)

	Week Ending Date	NEMSIS Calls	NEMSIS Calls CA Yes	NEMSIS Calls CA No	NEMSIS CA After EMS	NEMSIS CA Prior EMS	Influenza DX Codes	Respiratory Codes	COVID-19 DX Codes	COD COVID Primary	COD COVID Secondary	COD All Cause	Excess Deaths	Observed Deaths	Total Above
Week Ending Date	NA	ABOVE	BELOW	ABOVE	ABOVE	BELOW	ABOVE	ABOVE	ABOVE	BELOW	ABOVE	ABOVE	ABOVE	ABOVE	10
NEMSIS Calls	BELOW	NA	ABOVE	ABOVE	ABOVE	ABOVE	ABOVE	ABOVE	ABOVE	BELOW	BELOW	BELOW	BELOW	BELOW	7
NEMSIS Calls CA Yes	BELOW	ABOVE	NA	ABOVE	ABOVE	ABOVE	ABOVE	ABOVE	BELOW	BELOW	ABOVE	BELOW	ABOVE	BELOW	8
NEMSIS Calls CA No	BELOW	ABOVE	ABOVE	NA	ABOVE	ABOVE	ABOVE	ABOVE	ABOVE	BELOW	BELOW	BELOW	BELOW	BELOW	7
NEMSIS CA After EMS	BELOW	ABOVE	ABOVE	ABOVE	NA	ABOVE	ABOVE	ABOVE	BELOW	ABOVE	ABOVE	BELOW	ABOVE	BELOW	9
NEMSIS CA Prior EMS	BELOW	ABOVE	ABOVE	ABOVE	ABOVE	NA	ABOVE	ABOVE	ABOVE	ABOVE	ABOVE	BELOW	ABOVE	BELOW	10
Influenza DX Codes	BELOW	BELOW	BELOW	BELOW	BELOW	BELOW	NA	ABOVE	BELOW	BELOW	BELOW	ABOVE	BELOW	BELOW	2
Respiratory Codes	BELOW	BELOW	BELOW	BELOW	BELOW	BELOW	ABOVE	NA	BELOW	BELOW	BELOW	ABOVE	BELOW	BELOW	2
COVID-19 DX Codes	BELOW	BELOW	BELOW	BELOW	BELOW	BELOW	ABOVE	ABOVE	NA	BELOW	BELOW	BELOW	BELOW	ABOVE	3
COD COVID Primary	ABOVE	ABOVE	BELOW	ABOVE	BELOW	BELOW	ABOVE	ABOVE	ABOVE	NA	ABOVE	ABOVE	ABOVE	BELOW	9
COD COVID Secondary	ABOVE	BELOW	BELOW	BELOW	BELOW	BELOW	ABOVE	ABOVE	ABOVE	ABOVE	NA	ABOVE	BELOW	ABOVE	8
COD All Cause	ABOVE	BELOW	BELOW	BELOW	BELOW	BELOW	BELOW	ABOVE	BELOW	ABOVE	ABOVE	NA	ABOVE	ABOVE	7
Excess Deaths	BELOW	BELOW	BELOW	BELOW	BELOW	BELOW	ABOVE	ABOVE	ABOVE	ABOVE	ABOVE	ABOVE	NA	ABOVE	7
Observed Deaths	BELOW	BELOW	BELOW	BELOW	BELOW	BELOW	ABOVE	ABOVE	BELOW	ABOVE	ABOVE	ABOVE	ABOVE	NA	6

Abbreviations: COD, cause of death; COVID-19, coronavirus disease 2019; EMS Emergency Medical Services; NEMSIS, The National Emergency Medical Services Information System.

importance for "Week Ending Date," "COVID 19 DX Codes," and "COD COVID Primary" series. The "Total Above" ranged 5 to 8, indicating similar importance distributions.

In ►Table 4, the geometric mean has been computed for each row and if the raw value exceeds the geometric mean, the raw value is marked "above" as in ►Table 3. ►Table 4 can assess above average variable importance across models. High variable importance across models indicates that multiple series relied on the independent variable to learn the dependent weekly value. For example, in ►Table 4, "COD All Cause" independent variable was above the average variable importance (for different sources) models "Week Ending Date," "Influenza DX Codes," "Respiratory Codes," "Excess Deaths," and "Observed Deaths" (from excess deaths source). Total Above ranged from 2 to 10, suggesting that some series had acute agreement (small number) and some have generalized agreement. The Medicare sourced series have low Total Above, indicating their value is concentrated in models "COVID All Cause" and "Observed Deaths." Note that NEMSIS CA Prior EMS is tied with Week Ending Date in first place (10).

Discussion

Toward prior work, syndromic surveillance and the uses of prehospital data in understanding hospital utilization, (influenza) vaccination uptake, and community health are well described.[34–36] However, the potential for prehospital CA to be considered as a syndromic effect is perhaps limited to influenza and local area use cases in the United States.[37] The same cannot be said for Europe.[38,39] There is evidence that COVID-19 is associated with sudden cardiac death, some of which should be prehospital and pre-EMS arrival.[40] As influenza has inspired developments in syndromic surveillance, perhaps COVID-19 will do the same.[38]

Toward study findings, appreciating the severity of COVID-19 in the United States has been met with difficulty.[41–43] Preexisting surveillance methods have proven inadequate, and CDC has proposed a modernization effort to produce novel surveillance efforts within the epidemic response.[44] Ancillary events, such as EMS calls and Medicare bills, could support surveillance tasks like early detection of an outbreak, severity models, and prevention efforts. This paper demonstrates that Medicare and NEMSIS data have value when predicting traditional measures of epidemic modeling like COD and Excess Mortality.

Within Medicare sourced series, EMS call volumes were below average variable importance for Influenza and Respiratory Viral claims volumes but were above average for COVID-19 volumes when calls without CA and calls where CA occurred prior to EMS arrival are considered. NEMSIS series benefited from knowing the call volumes which were CA prior to EMS arrival, consistently ranked within NEMSIS series as 1 or the most important. COVID-19 as primary COD on a multiple COD certificate and the volume of Medicare COVID-19 claims was also above average in importance when predicting NEMSIS call volumes. This suggests that COVID-19 is driving EMS call volumes.

Within CDC MCDC series both primary and secondary COD models found above average predictive value from NEMSIS call volumes which involved a CA, suggesting that EMS arrests may not survive the experience. There is also predictive value in the CDC excess mortality model values but this is to be expected as the excess mortality model was designed to evaluate excess mortality from COVID-19. Within CDC Excess Mortality series, NEMSIS call volumes for CA as well as COVID-19 being present on a multiple COD certificate were high value when predicting the weekly Farrington Flexible mortality excess estimates.

Variable importance detailed in Tables 2 and 3 demonstrates meaningful model segmentation between series and series events. Influenza and viral respiratory codes are particularly interesting as a "control" case in this COVID-19 era data set. Both influenza and viral respiratory series show interrelatedness in their variable importance and difference or segmentation from COVID-19. "CA prior to EMS" arrival was also of note because "CA prior to EMS" arrival most likely results in a decedent without a COVID-19 diagnosis, a decedent who may be ineligible for a primary COD 'COVID-19' declaration. ►Table 3 further belabors the point, with "COD Primary COVID" model showing "NEMSIS Calls CA Yes," "NEMSIS CA Before," "NEMSIS CA Prior," "Observed Deaths," and "Excess Deaths" above the geometric mean of variable importance within the "COD Primary COVID" model. Given that DRF does not know what a cardiac arrest is nor Farrington Flexible but is still able to associate the weekly distributions with COVID-19 primary COD on MCDCs from only the weekly counts highlights the strength of this approach.

Table 4 demonstrates high general utility for most independent variables in the model series. It also suggests that the Medicare series was not as strongly utilized in decision-making with a geometric mean range of 2–3. This could be due to the real-world sampling distribution of Medicare enrollment relative to the total morbidity burden in the United States. How much of the COVID-19 burden should be among Medicare beneficiaries remains unknown. All other series are national, while Medicare is enrollee specific and may not offer as much instruction to prediction. However, despite the difference in real world lag (between claims being processed and a death certificate being populated, or a 911 call being placed), the model produced $r^2 > 0.9$ in most cases. Note that "NEMSIS CA Prior EMS" had as many "above" the geometric mean in ►Table 4 as the week itself. This means it is tied for the best predictor across models. The implications of these prior arrests are profound, and they may be a sink of underrecognized COVID-19 mortality.

The length of the series, and the "isotonic" nature of the data may explain the difficulty of predicting the week of series, as the opportunity for weekly patterns to repeat most likely confused week assignments. As COVID and influenza had multiple "waves" over the observation period, a bad week guess could be a repeat start, peak, or end event. A bad week guess could also be a time point with little data being confused for another low-volume time point. The NEMSIS anomaly in 2017 (low volumes) is not well understood but is most likely due to NEMSIS transitioning OLAP series in 2017

or perhaps there was a national decrease in EMS call volumes in 2017. Most likely the models are not impacted as the models consider records from 2018 onward.

The analysis would be more robust if series completeness could be achieved, especially in early model years. ►Table 1 shows several data series available in earlier years than others. Medicare data particularly suffers from changes in diagnostic code recall in ICD9-CM versus ICD10-CM years (only ICD10-CM years were considered here). The "stability" of a series is of high importance when evaluating future surveillance value. The model did not weigh variables by series source and did not "know'" that variables were from the same data sources. Weighting series completeness may improve model results; however, r^2 was high across models. The Medicare series contains diagnostic and pathology codes for influenza and COVID-19. There may be noncase incidence drivers of testing, vaccination, and pathology including nosocomial infections, the "worried well" as well as public health interventions (mass testing and roster vaccinations). Disambiguating the Medicare indexes could increase their utility even further. The viral respiratory code list includes minor codes like fever as well as ARDS and pneumonia. Their disambiguation by severity may improve model utility as well.

Conclusion

Prehospital data (EMS) are of high value in COVID-19 surveillance and should be considered as a potential data source when attempting to learn COVID-19 severity within jurisdictions. Medicare data faired weaker though individuals providing care to the Medicare population should consider the disambiguation of patients with COVID-19 from individuals seeking COVID-19 prevention services (testing and vaccination).

Human Subjects Protections

While this study contains identifiable information describing live human subjects, no National Institutes of Health Institutional Review Board (NIH IRB) review was required. Note that Centers for Medicare and Medicaid Services (CMS) data access and use are approved through the CMS IRB, however. Data were further "cleared" for public release by C.C.W., and C.C.W. evaluated our compliance with CMS nonreidentification standards for data describing beneficiary populations.

References

1 Le Duc JW, Sorvillo TE. A quarter century of emerging infectious diseases—where have we been and where are we going? Acta Med Acad 2018;47(01):117–130

2 Polonsky JA, Baidjoe A, Kamvar ZN, et al. Outbreak analytics: a developing data science for informing the response to emerging pathogens. Philos Trans R Soc Lond B Biol Sci 2019;374 (1776):20180276

3 Wu JT, Leung K, Lam TTY, et al. Nowcasting epidemics of novel pathogens: lessons from COVID-19. Nat Med 2021;27(03):388–395

4 Bhatia S, Lassmann B, Cohn E, et al. Using digital surveillance tools for near real-time mapping of the risk of infectious disease spread. NPJ Digit Med 2021;4(01):73

5 Leung K, Wu JT, Leung GM. Real-time tracking and prediction of COVID-19 infection using digital proxies of population mobility and mixing. Nat Commun 2021;12(01):1501

6 Redd SC, Frieden TR. CDC's evolving approach to emergency response. Health Secur 2017;15(01):41–52

7 Li BZ, Li MS, Huang JY, Chen YY, Lu YH. [Expanding the pandemic influenza preparedness framework to the epidemic of COVID-19]. Chin J Prev Med 2020;54(06):597–601

8 Tam T. Fifteen years post-SARS: key milestones in Canada's public health emergency response. Can Commun Dis Rep 2018;44(05): 98–101

9 Martinello RA. Preparing for avian influenza. Curr Opin Pediatr 2007;19(01):64–70

10 Gibson PJ, Theodore F, Jellison JB. The common ground preparedness framework: a comprehensive description of public health emergency preparedness. Am J Public Health 2012;102(04): 633–642

11 Brower JL. The threat and response to infectious diseases (Revised). Microb Ecol 2018;76(01):19–36

12 National Pandemic Influenza Plans | Pandemic Influenza (Flu) | CDC [Internet]. 2018 [cited 2022 Jan 18]. Accessed January 31, 2022 at: https://www.cdc.gov/flu/pandemic-resources/planning-preparedness/national-strategy-planning.html

13 Pandemic Influenza Preparedness, Response and Recovery Guide for Critical Infrastructure and Key Resources.:84.

14 Wang Q, Zhang T, Zhu H, et al. Characteristics of and public health emergency responses to COVID-19 and H1N1 outbreaks: a case-comparison study. Int J Environ Res Public Health 2020;17(12): E4409

15 Viglione G. How many people has the coronavirus killed? Nature 2020;585782322–24

16 Zhu H, Wang Q, Zhang T, et al. Initial public-health emergency response to SARS and COVID-19 pandemics in mainland china: a retrospective comparative study. Risk Manag Healthc Policy 2021;14:4199–4209

17 Lee H-Y, Oh M-N, Park Y-S, Chu C, Sona T-J. Public health emergency preparedness and response in Korea. J Korean Med Assoc 2017;60(04):296–299

18 Wang V. Why China Is the World's Last 'Zero Covid' Holdout. The New York Times [Internet]. 2021 Oct 28 [cited 2022 Jan 18]. Accessed January 01, 2023 at: https://www.nytimes.com/2021/10/27/world/asia/china-zero-covid-virus.html

19 Huang P, Ruwitch J. What the U.S. can learn from China's response to COVID infections. NPR [Internet]. 2021 Nov 8 [cited 2022 Jan 18]. Accessed January 01, 2023 at: https://www.npr.org/2021/11/08/1053647178/what-the-u-s-can-learn-from-chinas-response-to-covid-infections

20 Why China is still trying to achieve zero Covid. BBC News [Internet]. 2021 Nov 15 [cited 2022 Jan 18]. Accessed January 1, 2023 at: https://www.bbc.com/news/world-asia-china-59257496

21 Beaney T, Clarke JM, Jain V, et al. Excess mortality: the gold standard in measuring the impact of COVID-19 worldwide? J R Soc Med 2020;113(09):329–334

22 Iuliano AD, Chang HH, Patel NN, et al. Estimating under-recognized COVID-19 deaths, United States, March 2020-May 2021 using an excess mortality modelling approach. Lancet Reg Health Am 2021;1:100019

23 Rossen LM. Notes from the field: update on excess deaths associated with the COVID-19 Pandemic—United States, January 26,

2020–February 27, 2021. MMWR Morb Mortal Wkly Rep 2021; 70. Accessed January 1, 2023 at: cited 2021 Dec 15 [Internet] https://www.cdc.gov/mmwr/volumes/70/wr/mm7015a4.htm

24 Rossen LM. Excess Deaths Associated with COVID-19, by Age and Race and Ethnicity — United States, January 26–October 3, 2020. MMWR Morb Mortal Wkly Rep [Internet]. 2020 [cited 2022 Jan 11];69. Accessed January 1, 2023 at: https://www.cdc.gov/mmwr/volumes/69/wr/mm6942e2.htm

25 Stokes AC, Lundberg DJ, Elo IT, Hempstead K, Bor J, Preston SH. COVID-19 and excess mortality in the United States: a county-level analysis. PLoS Med 2021;18(05):e1003571

26 Woolf SH, Chapman DA, Sabo RT, Weinberger DM, Hill L. Excess deaths from COVID-19 and other causes, March-April 2020. JAMA 2020;324(05):510–513

27 Woolf SH, Chapman DA, Sabo RT, Weinberger DM, Hill L, Taylor DDH. Excess deaths from COVID-19 and other causes, March-July 2020. JAMA 2020;324(15):1562–1564

28 Farrington CP, Andrews NJ, Beale AD, Catchpole MA. A statistical algorithm for the early detection of outbreaks of infectious disease. J R Stat Soc Ser A Stat Soc 1996;159(03):547–563

29 Noufaily A, Enki DG, Farrington P, Garthwaite P, Andrews N, Charlett A. An improved algorithm for outbreak detection in multiple surveillance systems. Stat Med 2013;32(07):1206–1222

30 Dawson DE. National Emergency Medical Services Information System (NEMSIS). Prehosp Emerg Care 2006;10(03):314–316

31 Mann NC, Kane L, Dai M, Jacobson K. Description of the 2012 NEMSIS public-release research dataset. Prehosp Emerg Care 2015;19(02):232–240

32 Handberry M, Bull-Otterson L, Dai M, et al. Changes in Emergency Medical Services Before and during the COVID-19 pandemic in the United States, January 2018-December 2020. Clin Infect Dis 2021;73(Suppl (Suppl 1):S84–S91

33 Distributed Random Forest (DRF)—H2O 3.36.0.1 documentation [Internet]. [cited 2022 Jan 14]. Accessed January 1. 2023 at:

https://docs.h2o.ai/h2o/latest-stable/h2o-docs/data-science/drf.html

34 Onozuka D, Hagihara A. Extreme influenza epidemics and out-of-hospital cardiac arrest. Int J Cardiol 2018;263:158–162

35 Moa A, Tan T, Wei J, Hutchinson D, MacIntyre CR. Burden of influenza in adults with cardiac arrest admissions in Australia. Int J Cardiol 2022;361:109–115

36 Čulić V, AlTurki A, Proietti R. Public health impact of daily life triggers of sudden cardiac death: a systematic review and comparative risk assessment. Resuscitation 2021;162:154–162

37 Duijster JW, Doreleijers SDA, Pilot E, et al. Utility of emergency call centre, dispatch and ambulance data for syndromic surveillance of infectious diseases: a scoping review. Eur J Public Health 2020; 30(04):639–647

38 McVaney KE, Pepe PE, Maloney LM, et al; Writing group on behalf of the Metropolitan EMS Medical Directors Global Alliance. The relationship of large city out-of-hospital cardiac arrests and the prevalence of COVID-19. EClinicalMedicine 2021;34:100815

39 Riesgo LGC, Ziemann A, Rosenkoetter N, et al. Use of routinely collected emergency medical data for earlier detection of health threats in Europe: first evaluation results of the SIDARTHa syndromic surveillance system. Resuscitation 2010;81(02):S7

40 Yadav R, Bansal R, Budakoty S, Barwad P. COVID-19 and sudden cardiac death: a new potential risk. Indian Heart J 2020;72(05):333–336

41 Brodeur A, Gray D, Islam A, Bhuiyan S. A literature review of the economics of COVID-19. J Econ Surv 2021;35(04):1007–1044

42 Bergquist S, Otten T, Sarich N. COVID-19 pandemic in the United States. Health Policy Technol 2020;9(04):623–638

43 Xu HD, Basu R. How the United States flunked the COVID-19 Test: some observations and several lessons. Am Rev Public Adm 2020; 506–7568–576

44 Data Modernization Initiative [Internet]. [cited 2022 Jan 18]. Accessed January 31, 2023 at:2021https://www.cdc.gov/surveillance/data-modernization/index.html