International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 10 Issue 1 ǁ 2022 ǁ PP. 65-79

Algorithmic Prediction of Contending Risk Events (Hazards)

Achieng' Loice A
Agola Joshua
Liyala Samuel
[1]*Department of information systems and technology, Jaramogi Oginga Odinga University of Science and technology, Kenya*

Abstract
Healthcare emergency incidents are investable and often have long-term and short-term impacts on the well-being of an individual in the emergency care facility calling for the need to ensure that the emergency planning and response within the various healthcare facilities are effective. To mitigate such adverse risks from occurring, care providers must forecast the patients' potential risks and survival rates. While models such as Kaplan-Mier (KM) have been effective in the prediction of risk events in healthcare, it is ineffective considering it assumes that there would be an equal chance of observing the event of interest of the censored patient, which is not always the case as in many instances there are content risk events for in emergency care. The proposed model considers the contending risk events enabling the emergency care providers to prioritize the greatest risk in emergency care, thus reducing.

Keywords: Risk, Healthcare, Kaplan-Mier, Contending risk, Events.

Date of Submission: 15-01-2022 Date of acceptance: 30-01-2022

I. INTRODUCTION

Many healthcare emergency incidents have both long-term and short-term impacts on people's health while also placing additional demands on healthcare organizations [1]. The emergency planning and management to address these demands are often very complex. Besides, the many determinants of mental and physical well-being often imply that a wide variety of non-medical-care organizations may have roles in the prevention and recovery efforts, extending to the public, private, and voluntary sectors. Considering to dynamic nature of medical emergencies, they often come with competing risks. The concept of competing risk or competing hazards, particularly in mortality and healthcare victims of a medical emergency are likely to succumb to the injuries from the accident or suffer a stroke due to the shock from the accident mostly when only one of these can occur it is termed as "competing event." Thus, in emergency planning and response, healthcare service providers must conduct a competing risk analysis that can help estimate the marginal probability of an event in the presence of competing events.

II. Literature Survey

In medical situations and emergencies, the main outcome is the time to an event. Many studies evaluating time-to-event data often use the term "survival" to designate the methods and synonymously denote techniques used to estimate the occurrence of events at various follow-up periods for the population cohort [2]. The current body of research cites Kaplan-Meier (KM) as the most used method to evaluate time-to-event analysis. The model has received a lot of traction in academia because it is executable in most statistical packages. A study conducted by Sprung et al. in critical care applied this methodology to assess the influence of steroid therapy in patients with sepsis and septic shock on shock reversal [3]. The finding of the study confirmed that steroids decreased vasopressor dependence.

Nonetheless, different scholars and studies have criticized the use of KM in the case of competing events considering the methodology assumes that there would be an equal chance of observing the event of interest of censored patients [4]. This presumption is often incorrect, particularly in cases where there is no chance of the event of interest. This can be exemplified by a case where readmission to a hospital is an event of interest. The death of a patient serves as a competing event. When the patient dies, the risk of readmission ceases to exist regardless of the observation period. Thus, failing to consider the contending events in survival analysis results in overestimating the cumulative incidence of events [5]–[9]. Considering the pitfalls of the KM method, alternative methods must be designed or used for survival analysis.

III. Workings

Algorithmic Prediction of Contending Risk Events (Hazards)

Modeling regression for competing risk events

Fitting an AI regression model on competing risk events for a single risk Kaplan-Meier failure prediction, a survival function *S(t)* is determined such that the probability of a patient, or other objects of interest in healthcare emergency, surviving beyond any specified time-to-event is arrived at this way.

$S(t) = \Pr(\tau \geq t)$, where t = 0, 1, 2, … is the time

However, in most healthcare emergencies, specific causes of failure take place within the ambiance of several different facilitators in the presence of other different causes, which interfere with the probability of the singled-out event from occurring. These competing risk events such that even though a patient is exposed widely to other failure causes, the final failure can be accredited to a particular such failure. It was observed that the Kaplan-Meier analysis accurately estimated the likelihood of failure independent of any other competing cause, but the probability of one type of contending event was accurately approximated by utilizing the cumulative incident function (CIF). This CIF broke down the probability of an occurrence of failure into the likelihood of corresponding to each contending event such that in any given time, the comprehensive probability of failure was equated to the individual sums of CIFs in every type of competing risk occurrence. The researcher observed that CIF gave an approximate probability of a margin for a given occurrence as a function of its specific cause likelihood and the general survival likelihood.

The researcher noted that the CIF was an output of two approximates:

1) The approximation of hazardous occurrences at ordered failure time $\boldsymbol{t_f}$ for the occurrence-type of concern, denoted as:

$$\hat{h}_c(t_f) = m_{cf}/n_f$$

where the value of m_{cf} showed the number of occurrences for a given risk *c*, at a particular time $\boldsymbol{t_f}$ and with $\boldsymbol{n_f}$ as the total number of subjects.

2) The approximation of the general likelihood of surviving last time ($_{of-1}$):

$$\hat{S}(t_{f-1})$$

The researcher resolved that ***S(t)*** denoted the comprehensive survival function and necessarily not the cause specific survival function. It was realized that the reasonable grounds for taking the comprehensive survival function into key focus was for the simple reason that a patient would have made it through all other contending events for it to register a failure from the occurrence ***c*** at a time $\boldsymbol{t_f}$. Having the two approximates, the computation of the approximate incidence's likelihood of registering a failure from an occurrence ***c*** at a particular time $\boldsymbol{t_f}$ was noted as shown below:

$$\hat{I}_c(t_f) = \hat{S}(t_{f-1}) \times \hat{h}_c(t_f)$$

Therefore, it was observed that the likelihood of registering a failure from an occurrence ***c*** at a time was essentially the result of making it through the preceding periods and the specific cause of the hazardous event at a time $\boldsymbol{t_f}$. The researcher realized that the CIF for an occurrence ***c*** at a particular time $\boldsymbol{t_f}$ was resolved by calculating the total summation of time $\boldsymbol{t_f}$ (i.e., from $f'=1$ to $f'=f$) of the likelihood of the incidences overall occurrences c failures that were registered. The expression was then denoted as follows:

$$\mathbf{CIF}_c(t_f) = \sum_{f'=1}^{f} \hat{I}_c(t_f) = \sum_{f'=1}^{f} \hat{S}(t_{f'-1}) \times \hat{h}_c(t_{f'})$$

It was realized that CIF was the same as the 1-KM approximator in the absence of a contending event. In the presence of a contending event, the approximator 1-KM differed from the CIF because of the use of the function of survival ***S(t)*** that would count the number of failures from the contending occurrences plus the other occurrence of interest, where the approximator 1-KM was seen to utilize an occurrence specific survival function $\boldsymbol{S_c(t)}$, which treated the failures from the contending occurrences as censored. CIF bypasses the need to make unverifiable assumptions of independence of censoring on competing events by using the overall survival function. Since the *S(t)* is always less than $S_c(t)$, the CIF is always smaller than 1-KM estimates in competing event data, which means the 1-KM tends to overestimate the probability of failure from the event type of interest. Another advantage is that, by definition, the CIF of each competing event is a fraction of the *S(t)*; therefore, the sum of each hazard for all competing events should equal the overall hazard. This property of CIF makes it possible to dissect overall hazard, which has more practical interpretations.

It was observed that Klein and Zhang extensively covered different approaches to regression modeling with competing risks in the Handbook of Statistics, Survival Analysis (Klein JP, Zhang MJ, 2007) and by Moeschberger, Tordoff, Kocher in the Handbook of Statistics, a Review of Statistical Analyses for Competing Risks (Moeschberger ML, Tordoff KP, Kochar N, 2007). The researcher observed that the multivariable regression analysis in the presence of competing risk data became appropriate in studying the probabilities of

failure in a case where every single failure arises from at least one of the many causes, essentially looked at as contending occurrences. The study realized that the proportional Cox standards hazards model treated the contending risks of the occurrence of interest as a censored observation. Lunn and McNeil (Lunn M, McNeil D, 1995) proposed a regression model adapted from the Cox proportional hazard analysis with ***k***contending occurrences, such that each subject's data was duplicated ***k***times, a single instance for every failure variety, and the creation of ***k-1***variables for the verification of fulfillment of an occurrence. Non-proportional hazards can also be modeled through a stratified Cox regression.

The study observed that a semiparametric proportional hazard model that was advanced by Fine and Gray (Fine JP, Gray RJ, 1999) and by Klein and Andersen (Klein JP, Andersen PK, 2005) was one of the appropriate analyses to achieve a statistical significance for prognostic effects of pseudo values on the CIF for competing risks data that build upon the non-parametric test (Gray RJ, 1988). The contending risk analysis incorporated a non-parametric model, which utilized an adjusted Chi-squared test to yield a good comparison of the CIF curves among different groups. A parametric approach was noted to model the CIF based on the sub-distribution hazard function. This proportional hazard model modeled the CIF with covariates by treating the CIF curve as a sub-distribution function. The sub-distribution function is analogous to the Cox proportional hazard model, except that it models a hazard function derived from a CIF (known as a sub-distribution hazard). The Fine and Gray sub-distribution hazard function for event type c can be expressed as:

$$h_{c,CIF}(t) = \lim_{\Delta \to 0} \frac{Pr(t < T_c < t + \Delta t | Tc > t \cup T_{c'} \leqslant t, c' \neq c)}{\Delta t}$$

The above function estimates the hazard rate for event type ***c*** at time ***t*** based on the risk set that remains at time t after accounting for all previously occurring event types, including competing events.

It was then resolved that the CIF based proportional hazard model would then be defined:

$$h_{c,CIF}(t) = h_{0c,CIF}(exp)[\sum_{i=1}^{P} \gamma_i X_i]$$

This model satisfies the proportional hazard assumption for the subpopulation hazard being modeled, which means the general hazard ratio formula is essentially the same as for the Cox model, except a minor cosmetic difference that the betas in the Cox model is replaced by gammas in Fine and Gray's model. Consequently, we similarly interpret the gammas as we do for the betas estimated from a Cox model, except that it estimates the effect of certain covariates in the presence of competing events. The Fine and Gray model can also be extended to allow time-dependent covariates.

Fine and Gray (Fine JP, Gray RJ, 1999) defined a sub-distribution hazard of the CIF to represent the hazard of falling from a given cause in the presence of competing events, given that a subject has survived or has already failed due to different causes. In the above example of competing events, we might want to estimate, for example, the colon cancer mortality rate over time, and want to know whether the mortality rate of such colon cancer differs between two or more treatment groups, with or without adjustment of covariates, the typical approach involves the use of Kaplan Meier estimator to separately estimate the probability for each type of event, while treating the other competing events as censored in addition to those who are censored from loss to follow-up or withdrawal. This method of estimating event probability is called the cause-specific sub-distribution hazard function for cause**r**, which is expressed as

$$h_c(t) = \lim_{\Delta t \to 0} \frac{P(t \leqslant T_c < t + \Delta t | T_c \geqslant t)}{\Delta t}$$

The random variable $\boldsymbol{T_c}$ denotes the time to failure from event type ***c***, the cause-specific hazard function $\boldsymbol{h_c(t)}$ gives the instantaneous failure rate at time ***t*** from event type ***c***. The semiparametric proportional hazards model for the cause-specific subdistribution hazard of cause***r*** for a subject with covariate vector ***x*** is given by

$$h_c(t, X) = h_{0c}(t) exp(\sum_{i=1}^{p} \beta_{ic} X_i)$$

This proportional hazard model of event type ***c***, at time ***t***, allows the effects of the covariates to differ by event types, as the subscripted beta coefficient suggests.

$$\lambda_r(t|\boldsymbol{X}) = \lambda_{r0}(t) \exp(\beta_r^{\mathrm{T}} \boldsymbol{X})$$

Where $\boldsymbol{h_{0c}}$ is the baseline sub-distribution hazard of cause***r***, and $\boldsymbol{\beta_r^T}$ is the vector of coefficients for the covariates.

Modeling a mortality dataset with R

The researcher chose death as her subject. This section utilizes COVID-19 pandemic data. The study used a data collection of 17,412 weekly death reports from 1995 to 2021. However, it was noted that the findings were narrowed for 2019-2021. It was realized that the appropriate CIF procedures would accurately estimate each subject's failure rate for each mortality risk event. The regression model contoured respiratory difficulties like lung malfunctions that were plainly expressed in other well-known disorders like influenza that had high mortality among older generations. In this case, the researcher used data from the Dutch statistics institution CBS. This data was subsequently made available within R through the cbsodataR package.
The researcher obtained a mortality dataset sourced from the Open data from the Dutch, with access to the link as follows: Statistical Institute https://www.cbs.nl/en-gb/onze-diensten/open-data/statline-as-open-data.

Table 1: Mortality dataset sourced from the Open data within the Dutch Statistical Institute

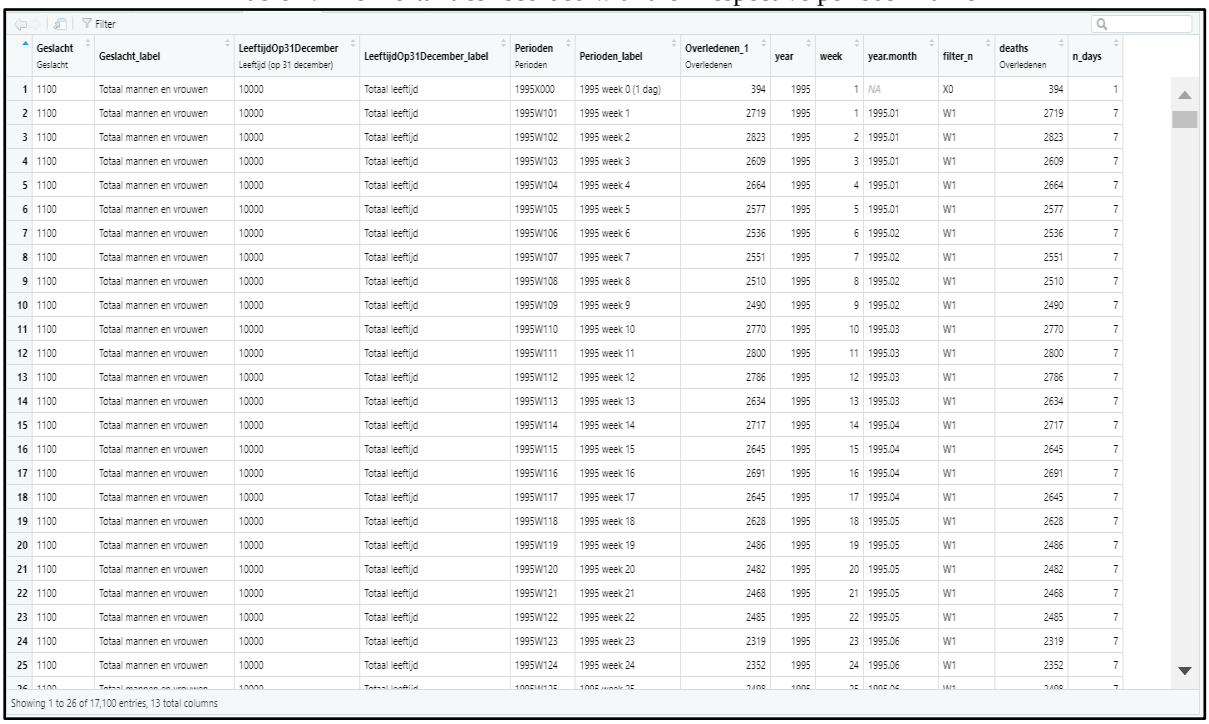
Filter

	Geslacht Geslacht	LeeftijdOp31December Leeftijd (op 31 december)	Perioden Perioden	Overledenen_1 Overledenen
1	1100	10000	1995X000	394
2	1100	10000	1995W101	2719
3	1100	10000	1995W102	2823
4	1100	10000	1995W103	2609
5	1100	10000	1995W104	2664
6	1100	10000	1995W105	2577
7	1100	10000	1995W106	2536
8	1100	10000	1995W107	2551
9	1100	10000	1995W108	2510
10	1100	10000	1995W109	2490
11	1100	10000	1995W110	2770
12	1100	10000	1995W111	2800
13	1100	10000	1995W112	2786
14	1100	10000	1995W113	2634
15	1100	10000	1995W114	2717
16	1100	10000	1995W115	2645
17	1100	10000	1995W116	2691
18	1100	10000	1995W117	2645
19	1100	10000	1995W118	2628
20	1100	10000	1995W119	2486
21	1100	10000	1995W120	2482
22	1100	10000	1995W121	2468
23	1100	10000	1995W122	2485
24	1100	10000	1995W123	2319
25	1100	10000	1995W124	2352
26	1100	10000	1995W125	2408

Showing 1 to 26 of 17,412 entries, 4 total columns

The study could approximate the COVID-19 death rate by using a cause-specific hazard model. To grasp risk analysis in emergencies, the research concluded. A random emergency occurrence point requires this information to analyze the circumstance. Various emergency points would have different consequences based on the severity of the crisis. Other causes of mortality, such as cardiac disorders, were also targeted. It was used to estimate COVID-19's emergency mortality consequences. This reflects well on the COVID-19 emergency points in the healthcare system. Using hazard analysis in emergency settings within or without healthcare facilities was useful. The researcher obtained the number of deaths documented in the dataset minus the projected number of deaths within a certain time. The researcher needed to estimate a typical mortality index. No COVID-19. A pandemic emergency case was identified. The death plans were then compared without the epidemic. A death statistics table based on historical trends was required to produce accurate forecasts within the study. They had to test statistical models against historical data. The researcher extended the data from the models to a future period. The researcher had to model historical data. Each had their take on the data and how it was used. One model was shown to be better than others in predicting outcomes. The researcher had to compare the models' prediction ability. This would enable a bias reduction technique for the AI model, as demonstrated below.

Table *2*: The mortalities recorded with their respective periods in time



	Geslacht	Geslacht_label	LeeftijdOp31December	LeeftijdOp31December_label	Perioden	Perioden_label	Overledenen_1	year	week	year.month	filter_n	deaths	n_days
	Geslacht		Leeftijd (op 31 december)		Perioden		Overledenen					Overledenen	
1	1100	Totaal mannen en vrouwen	10000	Totaal leeftijd	1995X000	1995 week 0 (1 dag)	394	1995	1	NA	X0	394	1
2	1100	Totaal mannen en vrouwen	10000	Totaal leeftijd	1995W101	1995 week 1	2719	1995	1	1995.01	W1	2719	7
3	1100	Totaal mannen en vrouwen	10000	Totaal leeftijd	1995W102	1995 week 2	2823	1995	2	1995.01	W1	2823	7
4	1100	Totaal mannen en vrouwen	10000	Totaal leeftijd	1995W103	1995 week 3	2609	1995	3	1995.01	W1	2609	7
5	1100	Totaal mannen en vrouwen	10000	Totaal leeftijd	1995W104	1995 week 4	2664	1995	4	1995.01	W1	2664	7
6	1100	Totaal mannen en vrouwen	10000	Totaal leeftijd	1995W105	1995 week 5	2577	1995	5	1995.01	W1	2577	7
7	1100	Totaal mannen en vrouwen	10000	Totaal leeftijd	1995W106	1995 week 6	2536	1995	6	1995.02	W1	2536	7
8	1100	Totaal mannen en vrouwen	10000	Totaal leeftijd	1995W107	1995 week 7	2551	1995	7	1995.02	W1	2551	7
9	1100	Totaal mannen en vrouwen	10000	Totaal leeftijd	1995W108	1995 week 8	2510	1995	8	1995.02	W1	2510	7
10	1100	Totaal mannen en vrouwen	10000	Totaal leeftijd	1995W109	1995 week 9	2490	1995	9	1995.02	W1	2490	7
11	1100	Totaal mannen en vrouwen	10000	Totaal leeftijd	1995W110	1995 week 10	2770	1995	10	1995.03	W1	2770	7
12	1100	Totaal mannen en vrouwen	10000	Totaal leeftijd	1995W111	1995 week 11	2800	1995	11	1995.03	W1	2800	7
13	1100	Totaal mannen en vrouwen	10000	Totaal leeftijd	1995W112	1995 week 12	2786	1995	12	1995.03	W1	2786	7
14	1100	Totaal mannen en vrouwen	10000	Totaal leeftijd	1995W113	1995 week 13	2634	1995	13	1995.03	W1	2634	7
15	1100	Totaal mannen en vrouwen	10000	Totaal leeftijd	1995W114	1995 week 14	2717	1995	14	1995.04	W1	2717	7
16	1100	Totaal mannen en vrouwen	10000	Totaal leeftijd	1995W115	1995 week 15	2645	1995	15	1995.04	W1	2645	7
17	1100	Totaal mannen en vrouwen	10000	Totaal leeftijd	1995W116	1995 week 16	2691	1995	16	1995.04	W1	2691	7
18	1100	Totaal mannen en vrouwen	10000	Totaal leeftijd	1995W117	1995 week 17	2645	1995	17	1995.04	W1	2645	7
19	1100	Totaal mannen en vrouwen	10000	Totaal leeftijd	1995W118	1995 week 18	2628	1995	18	1995.05	W1	2628	7
20	1100	Totaal mannen en vrouwen	10000	Totaal leeftijd	1995W119	1995 week 19	2486	1995	19	1995.05	W1	2486	7
21	1100	Totaal mannen en vrouwen	10000	Totaal leeftijd	1995W120	1995 week 20	2482	1995	20	1995.05	W1	2482	7
22	1100	Totaal mannen en vrouwen	10000	Totaal leeftijd	1995W121	1995 week 21	2468	1995	21	1995.05	W1	2468	7
23	1100	Totaal mannen en vrouwen	10000	Totaal leeftijd	1995W122	1995 week 22	2485	1995	22	1995.05	W1	2485	7
24	1100	Totaal mannen en vrouwen	10000	Totaal leeftijd	1995W123	1995 week 23	2319	1995	23	1995.06	W1	2319	7
25	1100	Totaal mannen en vrouwen	10000	Totaal leeftijd	1995W124	1995 week 24	2352	1995	24	1995.06	W1	2352	7

Showing 1 to 26 of 17,100 entries, 13 total columns

The researcher had to retrieve the population data from the same region. This enabled a proper indication of the nature of the population by gender and age groups. The following demonstrates the function calls for the procedure.
A sample extract of the results of the above process is shown in Table 6 below:

Table 3: Sample extract of the deaths per week per year

Year	Week	Deaths	year. month	n_days
1995	47	2613	1995.11	0
1995	48	2646	1995.11	0
1995	49	2826	1995.12	0
1995	50	3098	1995.12	0
1995	51	3291	1995.12	0
1995	52	3436	1995.12	0

The study had to integrate two different datasets: the mortality and population datasets. This facilitated the mathematical computation of the total number of mortalities per capita for every million subjects within the population.
Results are shown below:

Table 1: Total number of mortalities per capita for every million subjects within the population.

Men and women	2 years	2012	0.18569
Men and women	2 years	2013	0.184869
Men and women	2 years	2014	0.180172
Men and women	2 years	2015	0.176388
Men and women	2 years	2016	0.172395
Men and women	2 years	2017	0.176866
Men and women	2 years	2018	0.172815
Men and women	2 years	2019	0.174256
Men and women	3 years	1995	0.200037

Men and women	3 years	1996	0.197554
Men and women	3 years	1997	0.196222
Men and women	3 years	1998	0.197246

Table 4: Population data per year

2012	0.506544
2013	0.512131
2014	0.527228
2015	0.542534
2016	0.560345
2017	0.571885
2018	0.58874
2019	0.608726
2020	0.628029

The study moved ahead to do a filter application process function to arrive at the overall population as shown below:
The study demonstrated the details of the population dataset that was retrieved monthly as shown thus below:
Generated results were as shown in Table 9 below:

Table *5*: Generated results of the population dataset per month

17.18108	2018.01
17.18548	2018.02
17.19231	2018.03
17.1931	2018.04
17.1989	2018.05
17.20798	2018.06
17.21311	2018.07
17.2176	2018.08
17.23823	2018.09
17.2598	2018.1
17.27123	2018.11
17.28058	2018.12
17.28216	2019.01
17.29002	2019.02
17.29934	2019.03
17.3057	2019.04
17.31339	2019.05
17.32199	2019.06
17.32799	2019.07
17.33515	2019.08
17.35935	2019.09
17.38538	2019.1
17.39936	2019.11
17.40716	2019.12
17.40759	2020.01
17.41397	2020.02

17.42386	2020.03
17.42426	2020.04
17.41965	2020.05
17.42181	2020.06
17.4258	2020.07
17.43092	2020.08
17.44482	2020.09
17.4653	2020.1
17.47244	2020.11
17.47648	2020.12
17.47542	2021.01
17.47663	2021.02
17.48202	2021.03
17.48821	2021.04
17.49469	2021.05
17.50052	2021.06
17.50713	2021.07
17.51515	2021.08
17.53552	2021.09

Table 6: Data extract from population

2020	48	3405	2020.11	0	17.38337	0.812647	2.581335	17.47244	195.8768	4.674852	14.84945	19.5243	11
2020	49	3530	2020.12	0	17.38337	0.812647	2.581335	17.47648	203.0676	4.674852	14.84945	19.5243	11
2020	49	3530	2020.12	0	17.38337	0.812647	2.581335	17.46964	203.0676	4.674852	14.84945	19.5243	11
2020	50	3615	2020.12	0	17.38337	0.812647	2.581335	17.47648	207.9573	4.674852	14.84945	19.5243	11
2020	50	3615	2020.12	0	17.38337	0.812647	2.581335	17.46964	207.9573	4.674852	14.84945	19.5243	11
2020	51	3909	2020.12	0	17.38337	0.812647	2.581335	17.47648	224.8701	4.674852	14.84945	19.5243	11
2020	51	3909	2020.12	0	17.38337	0.812647	2.581335	17.46964	224.8701	4.674852	14.84945	19.5243	11
2020	52	6236	2020.12	4	17.38337	0.812647	2.581335	17.47648	358.7336	4.674852	14.84945	19.5243	11
2020	52	6236	2020.12	4	17.38337	0.812647	2.581335	17.46964	358.7336	4.674852	14.84945	19.5243	11

The data from mortality and population was later merged as shown below:
Generated results from the plot:
The researcher had to go ahead and plot the average mortality rate per week against the total number of mortalities in every given year, which showed the tendencies that were there before the year 2007 and those that were observed after the year 2010 significantly differed. This was useful since the inclusion of a yearly timing pattern in the predictions made together with the older patterns before the year 2007 was seen to have the potential to distort the predictions.

Modeling process with the merged dataset

The researcher then started modeling the two datasets. Initially, the researcher had to create some basic models. These were built to compare the most complicated models. The researcher employed basic ways to test the model's accuracy. The researcher had to clear the database of all 2020 and 2019 records. Incorporating data into models that solely represented information before 2018 was required. We used 2019 data. The researcher must next compare the deduced work to the 2019 data. Due to the nature of the data assessed, non-sampled forecasts were considered. Comparing the model's study predictions to actual data, i.e., data that wasn't utilized to approximate the model parameters, revealed a good resolution.
. The results are shown below:

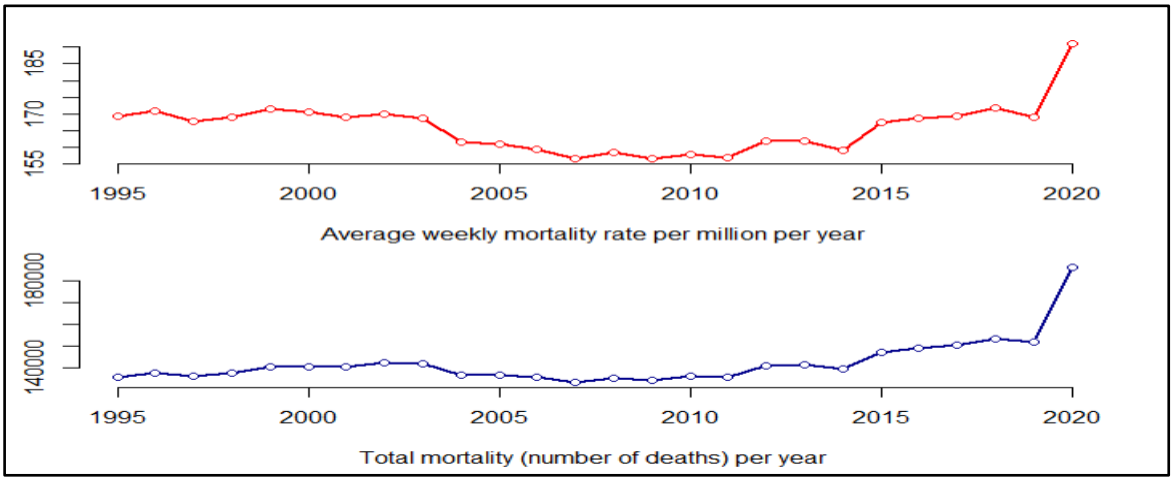


Figure 1: Model prediction results

The modeling process continued with the creation of two separate data subsets that were for different functions. One set was used for estimation, while the other was utilized in the evaluation process.

To compare how well the models forecast, two measures were required: the mean absolute prediction error (MAE) and the mean squared prediction error (MSE) (MSE). The study found that MAE took the absolute values for the discrepancies, whereas MSE took the squared values. The research investigation found MAE valuable in that it could directly test the accuracy of the mortality estimates. It was noted that these either came out too frequently or too infrequently. However, the utility of MSE was proved by its ability to obtain error details that were closely connected to the samples used. The study found that MSE penalized the heavier prediction mistakes the most since they were squared.

Linear regression model (LRM) fitting

The researcher used the initial model's linear regressive characteristics to represent a week as a categorical variable. The equivalent of this was using weekly averages from 2010 to 2018 to estimate the same week in 2019. The LRM was manufactured in two variants. The dependent variable was the mortality rate, whereas the dependent variable in another was the raw mortality numbers. The results indicated that every week was treated as a distinct class, with its regression coefficient. The model also can provide weekly average forecasts. Using this precise data, the team obtained the model's 2019 forecasts and correct process flaws. The study then ran a model utilizing raw mortality data as the dependent variable, yielding intriguing results.

The model's MAE was 6.34, with a mortality rate as the dependent variable. Multiplying this value by the number of weeks in a year and the total population for the census count of 2019 found 5698 predicted deaths. This was deemed incorrect. Similarly, the other comparative research model employed real raw mortalities as the dependent variable and predicted actual mortality with an MAE of 164.84. The study concluded that the MSE errors could not be easily compared since the dependent variables were of different scales. The study concluded that the two models might be utilized to create standards despite their simplicity.

The model's MAE was 6.34, with a mortality rate as the dependent variable. Multiplying this value by the number of weeks in a year and the total population for the census count of 2019 found 5698 predicted deaths. This was deemed incorrect. Similarly, the other comparative research model employed real raw mortalities as the dependent variable and predicted actual mortality with an MAE of 164.84. The study concluded that the MSE errors could not be easily compared since the dependent variables were of different scales. The study concluded that the two models might be utilized to create standards despite their simplicity. It was seen that the two models gave a poor performance than the ones that had a more simplified trend for linear time. It was noted that the entire population's coefficients within the country turned out positive and were quite remarkable. However, they could not give a much wider predictive leverage compared to the linear time trends. The variables were seen to correlate with one another highly. However, it was observed that not all were very relevant for accurate predictions except one.

The best model employed people between 65 and 80 as predictors. The lead analysis closely followed the percentage of those aged 80+. The investigation found no significant differences in the models' ability to forecast. The chart below compares the model's performance. The model predicts the temporal trend in red, whereas the 65-80 age group shares a green.

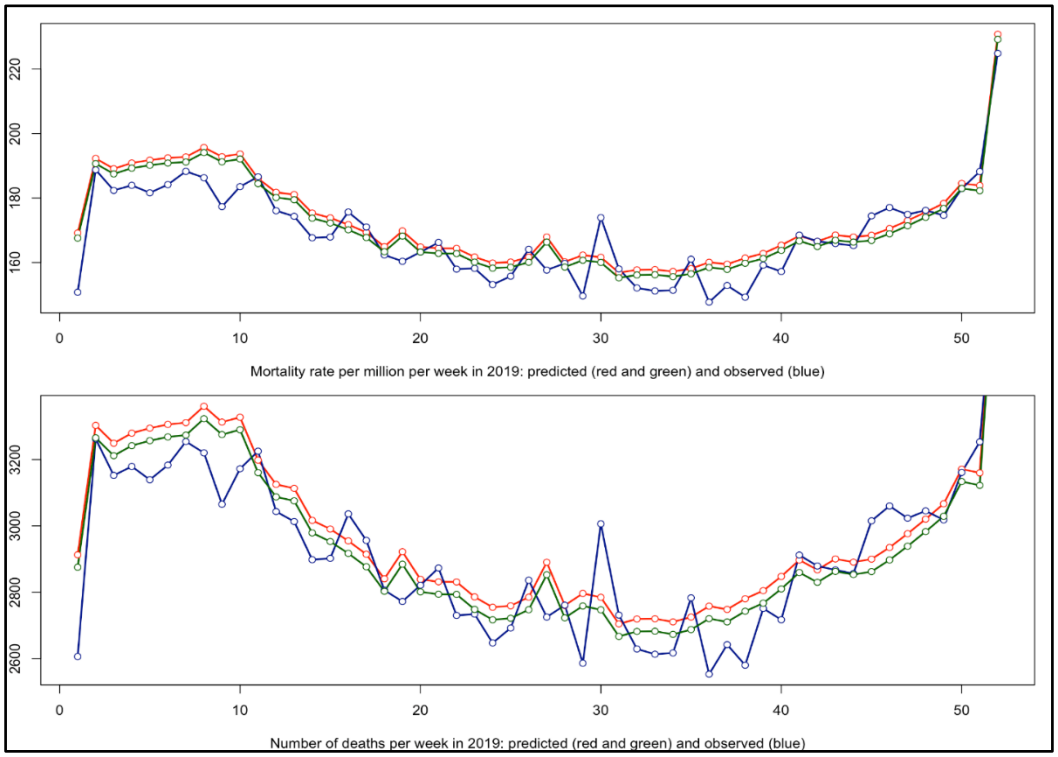


Figure 2: Models predictive performance

The study found that the models overstated the predicted outcome up to week 40 and after that had a very close relationship to the observed numbers. After a few weeks of adequate indicators for time trends, a dramatic shift was detected up to the 40th week. As high temperatures were shown to affect death rates positively, it was required to add weather information to the predictors before continuing with the research.

A characteristic that cut across the population's variables showed just one change every year. This was not practical since population dynamics occur over time. The scientists next tested if adding monthly population statistics would help. The study found worsening forecasts. The optimal model would thus number the raw mortalities as a weekly function, considering the older age group.

The risk event time lag

The work improved the model fitting by adding delayed death rates and death rates. Autocorrelation (ACF) and partial autocorrelation (PACF) functions were substantially established for the first order and preceding order autocorrelation. This meant that mortality rates for the current week were highly associated with death rates for two weeks ahead.

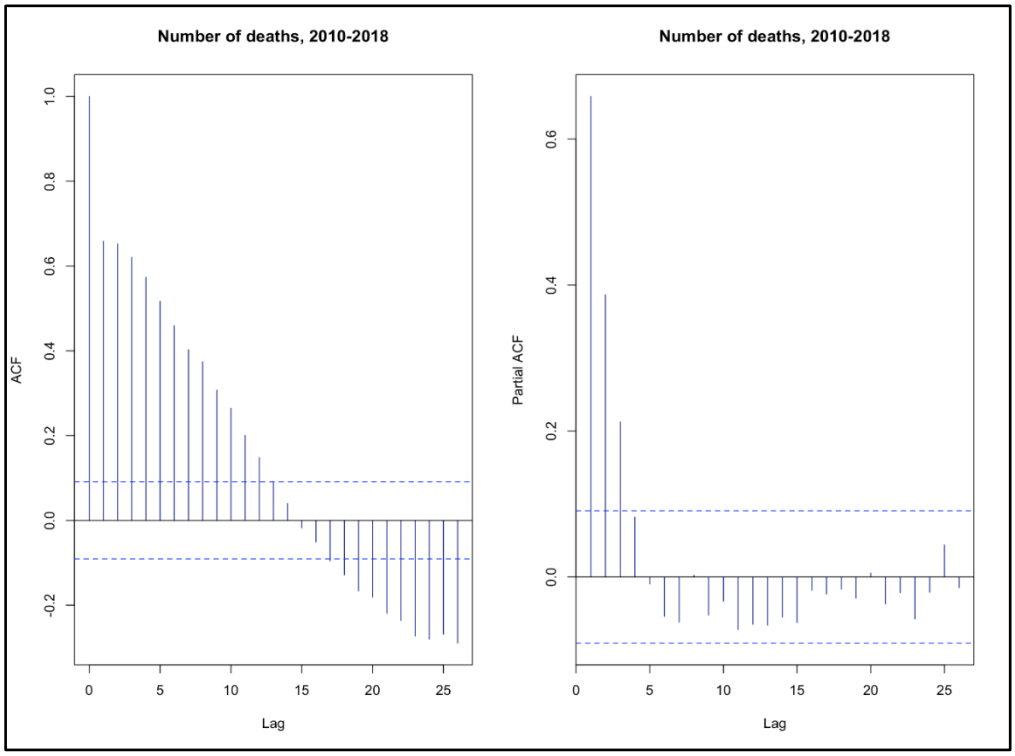


Figure 3: Comparative analysis on number of deaths

Adding the first and second lags greatly improved the model's fit. It let the researcher estimate fatalities one and two weeks in the future with a lag model. Each week was a category variable. Every week had one. In a realistic scenario, there should be weekly modifications. Less over-fitting was predicted because of the weekly dummies. A sine curve might substitute a weekly dummy for a more flexible period. The model's prediction performance was superior to the standard. It became a reserve option. Figure 4 displays the correct sine curve model forecasts in red, the data for 2019 in blue, and the weekly dummies predictions in green.

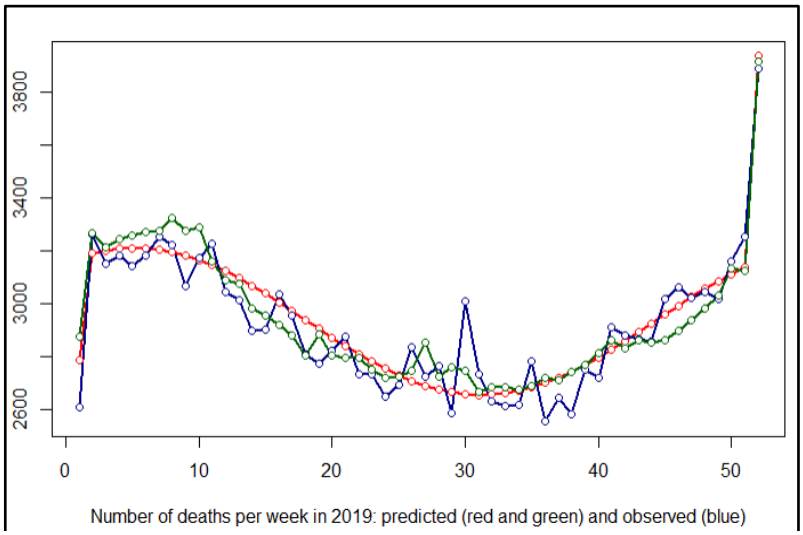


Figure 4: Predictions from the sine curve model

Fitting a Non-Linear model

During the investigation, it was noted that all prior models were linear. However, the weekly death rate and actual mortalities surpassed with significant frequency. These numbers deviated greatly from the normal distribution curve. The findings are shown in the histograms in figure 5:

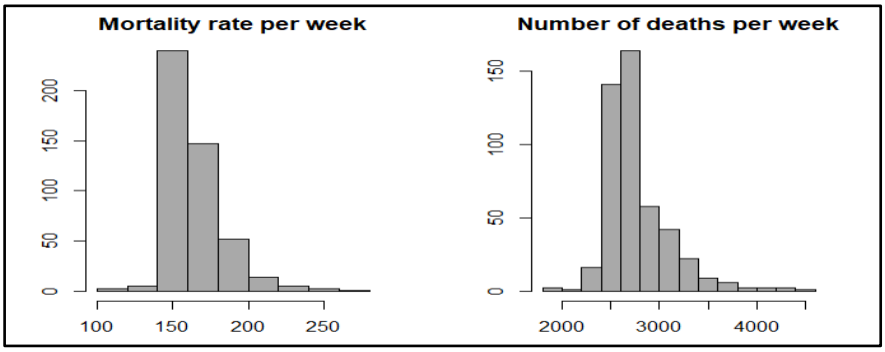


Figure 5: Histogram showing mortality rate per week as well as deaths per week

The researcher examined generalized linear models for dependent variables that normally follow a distribution in response. The investigation required negative binomial model fitting using MASS R's package. This usage was required due to its capacity to tolerate an overdispersion.

The study found that the model's ability to anticipate data from 2019 reliably was quite good. But it wasn't as good as the standard. The study concluded that the negative binomial model should be retained. The explanation was that the negative binomial model performed better for over-dispersed data counts than the linear model. The research found that producing approximation values for uncertain predictions was difficult, especially using the negative binomial model.

Robust regression modeling

The analysis found that specific weeks from 2010 to 2018 had very high fatality rates. This was due to competing risks such as flu outbreaks and heatwaves. These striking findings seem to impose disproportionate impacts on the models. This would result in model inclinations that would have varying habits from the start and predictable consequences. The investigation found three main outliers in the next model. These influenced the model, especially figure 6 below.

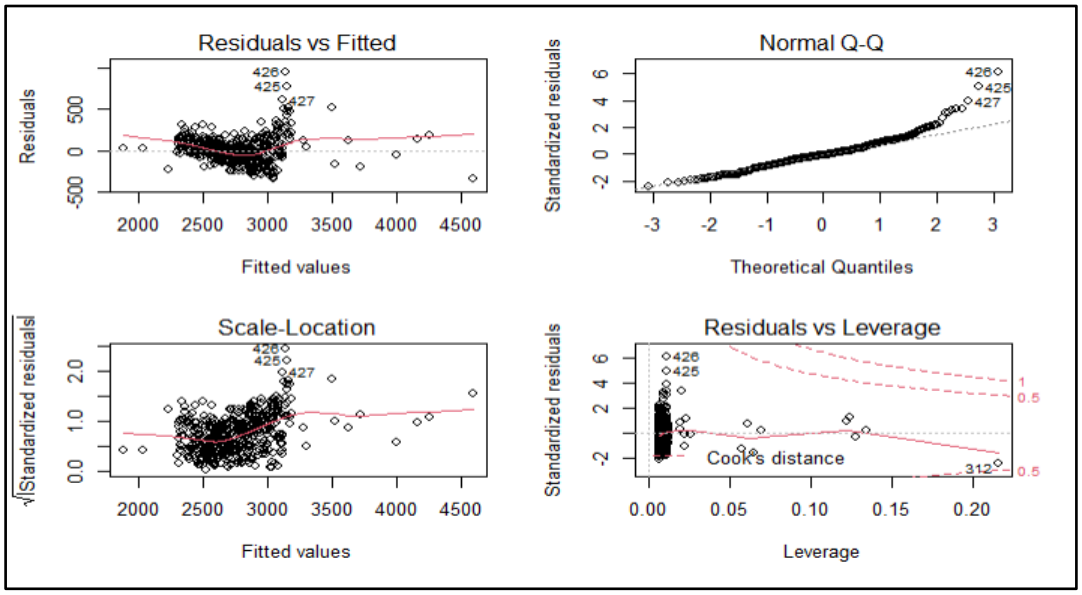


Figure 6: robust regression model outcomes

The study found that robust regression was a superior technique to removing the three observations. The study discovered that robust regression might reweight the data. This was based on the nature of the unexpected. The investigation found that the outliers suspected as an outcome were not eliminated from the data utilized and contributed little to the model's approximates.

The model had the best prediction ability, with an MAE of 72.06. Compared to the standard, it outperformed it. Figure 7 compares the robust model's predictions to those of the non-robust model.

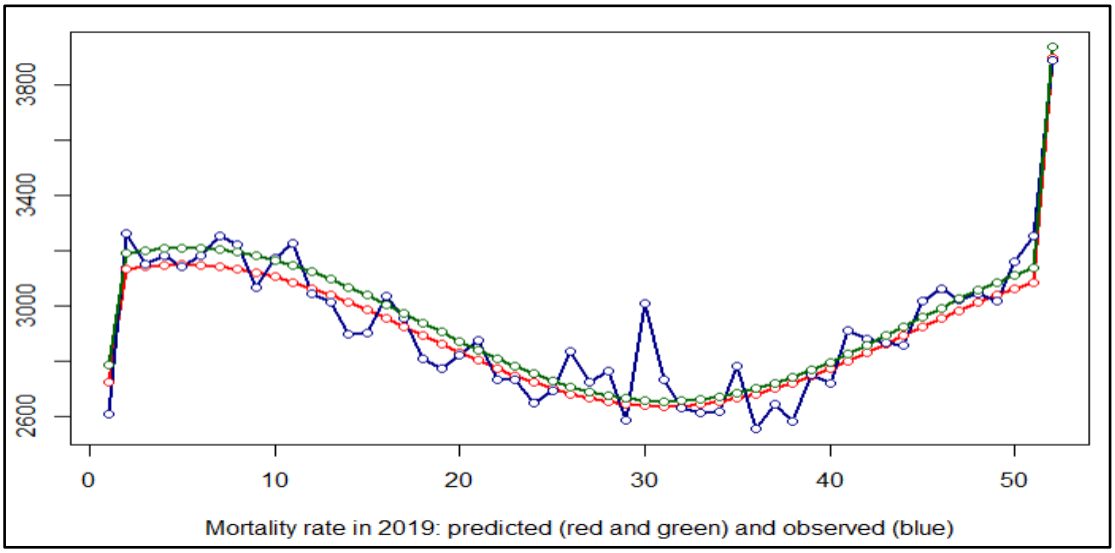


Figure 7: Comparative prediction outcomes of the robust and non-robust model

Adding temperature risk event

The analysis concluded that some data outliers were due to severe weather conditions, with some being extremely low and others very high. The researcher then added temperature data to the model to see whether any significant differences were found. Notably, the rnoaa program was integrated, which generally provides global weather data. Rnoaa was used to gather daily minimum and maximum temperatures from a Dutch weather station. The researcher produced daily and weekly aggregate numbers for the mortality and population statistics. The research then attached the newly discovered meteorological data with the remainder. The study predicted temperature not to affect mortality, yet it did. The study used a residuals graphic from the models compared to weekly temperature records for the lowest and maximum values to clarify the threshold points.

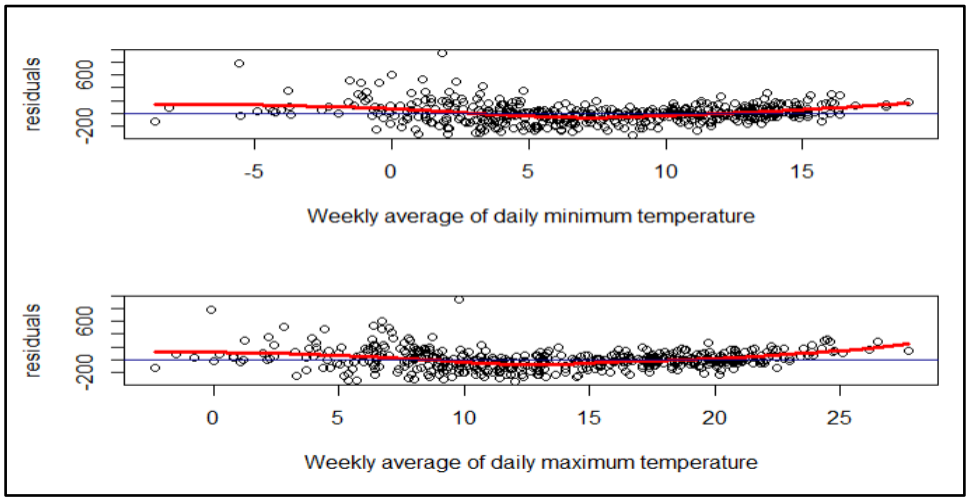


Figure 8: Residual plots to observe the point of increase

Extreme weather might explain certain data outliers, such as cold or hot periods. We can add temperature data to the model to see whether it matters. To view open historical weather data from around the world, install rNOAA. We'll extract a time series of daily minimum and maximum temperatures from a single Dutch weather station and then aggregate them to match mortality and population numbers.

Now we can link the new weather data to the rest of the data.

It's up to us to model temperature. Temperatures in the usual range should not influence mortality, yet they do. We may plot the residuals from one of our models (Model 8b) against the weekly minimum and maximum temperatures to find these boundaries.

Uncertainty and prediction intervals

After reviewing the various models, the two finalists had to be chosen. One model specified resilient linearly regressive forms, whereas the other specified negatively binomial functions. Both models used raw mortality numbers as the dependent variable, and the sine trend, 65-80 age bracket share, temperature, and weekday count as predictor functions. The researcher then went on to calculate the genuine prediction uncertainty.

Prediction intervals can be used to assess model uncertainty. The study avoided any misunderstanding with confidence intervals (which represented uncertainty about the anticipated value of a result as a function of some predictor value). The researcher might readily get the robust linear regression model prediction intervals by using a prediction function.

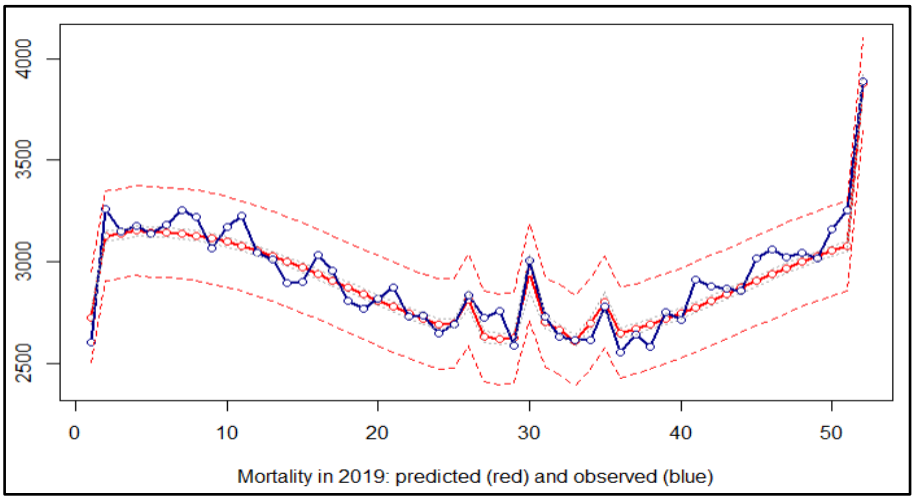


Figure 9: Mortality in 2019 predicted (red) and observed (blue)

As seen in the graph, the prediction intervals were rather large. They must have been too gentle because no value reported in 2019 dropped outside the cognizable intervals. The gray lines represented the degree of confidence and were solely used for demonstration reasons; they were much thinner and hence ineffective for evaluating the prediction uncertainty outcomes. So the prediction interval coverage has to be compared to the model estimations.

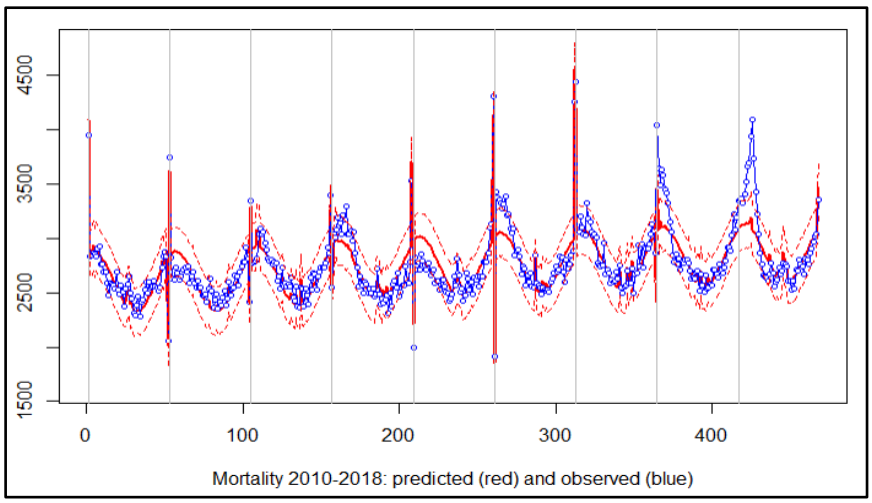


Figure 10: Mortality 2010-2018 predicted (red) and observed (blue)

The prediction intervals were sufficiently covered to provide a dataset for 'training' that included values outside the 95 percent interval predictions. Aside from seasonal cycles and minimum temperature, these were strongly tied to pandemic flu peaks. The study found that anticipating them was difficult and that the prediction intervals were too wide. The researcher then addressed the negative binomial specification, which was intended to manage the dependent variable's overdispersion better. The image below shows no substantial differences between the inappropriately behaved weeks from week one to week 52.

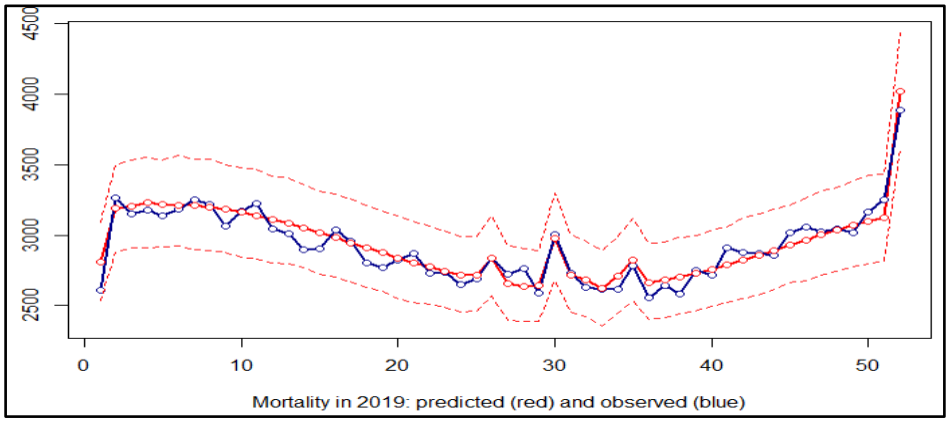


Figure 11: Inappropriately behaved weeks ranging from week one to week fifty-two

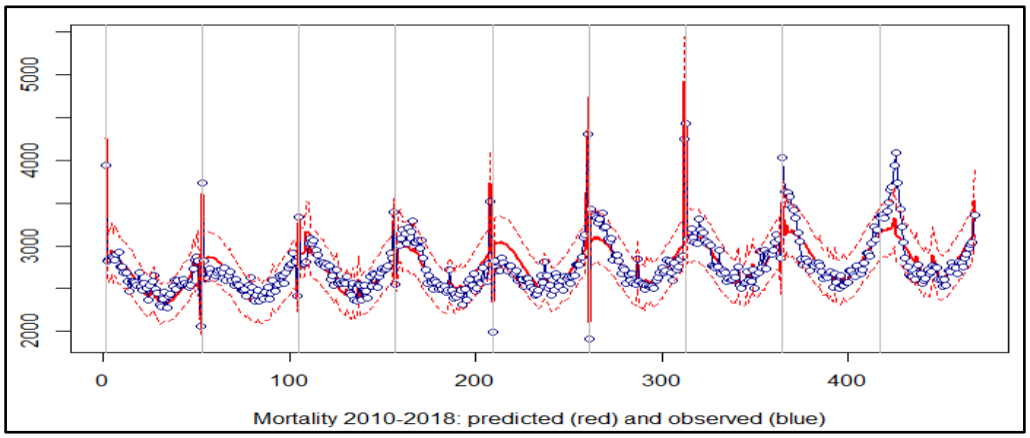


Figure 8: Mortality predictions and observations

Using these models to predict mortality

Finally, the researcher resolved that the algorithms would be appropriate in forecasting mortality data for 2020. Our models were seen to be robust linear regression and negative binomial regression. It was, therefore, necessary to assess their prediction uncertainty. Mispredictions resulted in 322687 fatalities, while the MSE was 366129. These indicators are much higher than any year between 2015 and 2019. In 2018, we had the worst predictions ever. Based on models that predicted mortality for any year between 2015 and 2019, fatality in 2020 appears to be exceedingly rare and unpredictable. The researcher then went ahead to plot the results. Death rates in 2020 were shown in blue (as measured by CBS). The negative binomial model's predictions are presented in green, whereas the robust linear regression model's predictions are red. The negative binomial model's prediction intervals were slightly larger, but the two models are otherwise identical. Remember that the 95 percent prediction intervals were wide and that none of the observed values for 2019 approached the prediction intervals' limits.

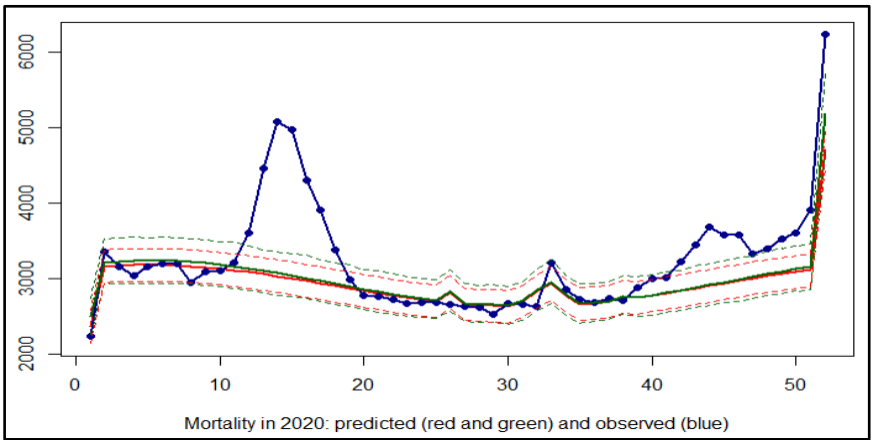


Figure 9: Mortality in 2020 predicted (red and green) and observed (blue)

To be precise, additional mortality in 2020 was 15613 fatalities, based on the best forecasts available. This represents 10.21% of the predicted total number of fatalities for the year. COVID-19 could not easily be held accountable for all additional mortality. There have been no wars or natural disasters this year, no evidence of a significant flu epidemic, and the model appears to have adequately predicted the summer heat-related mortality, so it has already been incorporated in. As a result, COVID-19 is responsible for the observed increase in mortality. Returning to the primary objective of this modeling exercise, we can state that our best estimate for the impact of COVID-19 on mortality, based on 2020 data, was an increase of 15613 deaths.

IV. CONCLUSION

The contending risk model helps enhance emergency case response in healthcare by predicting the most likely health hazard and directing resources towards mitigating such risks. The model determines the probability of a particular risk occurring, especially when risks would mean that the alternative risk is avoided. The model is important in the emergency health response as it allows the emergency service providers to prioritize the greatest risks that are likely to give the best forecast of fatality with respect to the contending risk events in the emergency healthcare setting. In this respect, it is important for further research to be conducted on contending risk and how best the model can be utilized to help in mitigating risk by forecasting the most likely risk event.

REFERENCES

[1] A. Boyd, N. Chambers, S. French, D. Shaw, R. King, and A. Whitehead, "Emergency planning and management in health care: priority research topics," *Health Syst (Basingstoke)*, vol. 3, no. 2, pp. 83–92, Jun. 2014, DOI: 10.1057/hs.2013.15.

[2] R. B. Moraes, G. Friedman, T. Lisboa, M. V. Viana, V. Hirakata, and M. A. Czepielewski, "Comparison of cumulative incidence analysis and Kaplan-Meier for analysis of shock reversal in patients with septic shock," *Journal of Critical Care*, vol. 27, no. 3, p. 317.e7-317.e11, Jun. 2012, DOI: 10.1016/j.jcrc.2011.06.006.

[3] C. L. Sprung *et al.*, "Hydrocortisone therapy for patients with septic shock," *New England Journal of Medicine*, vol. 358, no. 2, pp. 111–124, 2008.

[4] M. W. Donoghoe and V. Gebski, "The importance of censoring in competing risks analysis of the subdistribution hazard," *BMC Medical Research Methodology*, vol. 17, no. 1, p. 52, Apr. 2017, DOI: 10.1186/s12874-017-0327-3.

[5] J. M. Satagopan, L. Ben-Porat, M. Berwick, M. Robson, D. Kutler, and A. D. Auerbach, "A note on competing risks in survival data analysis," *British journal of cancer*, vol. 91, no. 7, pp. 1229–1235, 2004.

[6] J. C. Keurentjes, M. Fiocco, B. W. Schreurs, B. G. Pijls, K. A. Nouta, and R. Nelissen, "Revision surgery is overestimated in hip replacement," *Bone & joint research*, vol. 1, no. 10, pp. 258–262, 2012.

[7] W. D. Leslie, L. M. Lix, and X. Wu, "Competing mortality and fracture risk assessment," *Osteoporosis International*, vol. 24, no. 2, pp. 681–688, 2013.

[8] S. D. Berry, L. Ngo, E. J. Samelson, and D. P. Kiel, "Competing risk of death: an important consideration in studies of older adults," *Journal of the American Geriatrics Society*, vol. 58, no. 4, pp. 783–787, 2010.

[9] P. C. Austin, D. S. Lee, and J. P. Fine, "Introduction to the analysis of survival data in the presence of competing risks," *Circulation*, vol. 133, no. 6, pp. 601–609, 2016.