

Open Access

Time Series Analysis on Diabetes Mortality in the United States, 1999-2015 by Kolmogorov-Zurbenko Filter

Stella Arndorfer* and Igor Zurbenko

School of Public Health, State University of New York, Albany, Rensselaer, NY 12144, USA

Abstract

Kolmogorov-Zurbenko filters can be utilized in the public health context analyzing mortality data. This paper aims to expand upon the robust methodology of the KZ filters and their many applications. As a low-pass filter the KZ filters are proven to be the optimal means of analysis for non-stationary data such as mortality data which usually contains various underlying signals: seasonality, long-term trend, and short-term fluctuations. As diabetes incidence and prevalence increases, the burden of health care cost increases, thus prompting the need to understand patterns underlying adverse events related to diabetes, such as mortality. Increasing incidence and prevalence of diabetes prompts the need for preventative measures and understanding what environmental factors are related to adverse events as a result of diabetes. Diabetes mortality across time analyzed with non-parametric models has not previously been studied, thus this extension to the KZ filters is utilized as a preliminary analysis to address the gap in knowledge of diabetes mortality in the United States. Non-parametric time series analysis methods identify an 8.5-year long-term trend as well as annual seasonality of diabetes mortality. Spectral and time analysis of diabetes mortality introduces the relationship between solar activity and diabetes mortality which is quantified utilizing the cross-correlation between diabetes mortality and total solar irradiation. The strong correlation between solar activity and diabetes mortality.

Keywords: Time series; Kolmogorov-Zurbenko filters; Diabetes mortality; Total solar irradiation

Introduction

Kolmogorov-Zurbenko (KZ) filters are commonly used in the analysis of time series data because they do not rely on parametric assumptions [1,2]. KZ filters are classified as low-pass filters, which have the ability to decompose the original signal into its underlying separate components. Isolation of these varying components is especially important for signals that are suppressed due to noise, non-linearity and non-stationarity [1-5]. In this paper, KZ filters are utilized on diabetes mortality data in the United States from January 1999-December 2015.

As diabetes prevalence and the cost of diabetes medical care costs continue to increase it becomes more urgent to understand how diabetes has manifested across time and space [6,7]. Current literature that conduct time series analysis diabetes employ classic time series methods [8-11]. These classic time series data are parametric and do not accurately depict the original due to violations of the associated assumptions. Parametric time series methods are not able to accurately account for the complexity of non-stationary and non-normally distributed data [1,2,12-15]. Much like most mortality data, diabetes mortality time series displayed a non-stationary signal indicating the necessary use of non-parametric methods such as the KZ Filter methodology [8-11].

To our knowledge this is the first analysis employing nonparametric methods to understand the complex nature of diabetes mortality in the United States. Landmark studies in diabetes such as the UK Prospective Diabetes Study (UKPDS), Diabetes Control and Complications Trial (DCCT), Action to Control Cardiovascular Risk (ACCORD), Action in Diabetes and Vascular Disease (ADVANCE), Diabetes Prevention Program (DPP) all greatly contributed to current knowledge in risk factors, prevention measures and outcomes related to diabetes [16-23]. In addition to these landmark studies, many supplementary studies that examine diabetes, mortality and adverse events related to diabetes use multiple measures to examine their relationship to adverse diabetes outcomes such as Hemoglobin A1c (HbA1c), micro/macro-albuminuria, extended Globular Filtration Rate (eGFR) and demographic risk factors [8-11,16-24]. Tseng et al. found that HbA1c has seasonal variations similar to patterns found in mortality [9]. Thus, landmark trials and supplementary trials may benefit from a time series analysis that proves the existence of seasonal differences in diabetes mortality that may be related to seasonal fluctuations in HbA1c. Future studies may benefit from this analysis to control for seasonal differences in winter compared to summer months.

KZ Filters are especially beneficial to accurately predict future trends by decomposing the original signal and re-constructing it with high accuracy in the presence of non-stationary and noisy data [1,2]. Furthermore, this paper introduces the correlation between diabetes mortality and solar radiation measures in time. This relationship is quantified due to the discovery of corresponding frequencies found in spectra of radiation and sun activity and the diabetes time series spectra [1,2,25,26]. To our knowledge, this is the first analysis to quantify the relationship between radiation and diabetes mortality. In the presence of high rates of mortality, especially among vulnerable populations, it is important to understand environmental exposures that affect diabetes adverse effects.

Received Noveber 03, 2017; Accepted December 20, 2017; Published December 29, 2017

Citation: Arndorfer S, Zurbenko I (2017) Time Series Analysis on Diabetes Mortality in the United States, 1999-2015 by Kolmogorov-Zurbenko Filter. J Biom Biostat 8: 384. doi: 10.4172/2155-6180.1000384

Copyright: © 2017 Arndorfer S, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

^{*}Corresponding author: Stella Arndorfer, School of Public Health, State University of New York, Albany, Rensselaer, NY 12144, USA, Tel: 515-320-6296; E-mail: sarndorfer@albany.edu

Background

In 2015, diabetes was reported as the 7th leading cause of death in the United States and is known to lead to serious complications such as: kidney failure, cardiovascular disease, retinopathy, and neuropathy [27]. The National Diabetes Statistics Report reports approximately 30.3 million people have diabetes, which accounts for 9.4% of the United States population in 2015. Of these 30.3 million with diabetes 7.2 million (23.8%) were undiagnosed. Currently 1 in 11 individuals will develop diabetes; this number dramatically increases for ethnic minorities such as African Americans and Hispanics. The increasing prevalence of diabetes and prediabetes in the United States has incurred heavy economic burden for those diagnosed and for the United States health care system [27,28].

In 2017, prediabetes diagnoses have been recommended by the Standards of Medical Care in Diabetes in order to decrease the number of undiagnosed cases and to implement preventative interventions among those diagnosed with prediabetes [24]. These new diagnoses will likely increase the number of diabetes diagnoses in future years. According to the CDC, 84.1 million under the age of 18 were diagnosed with pre-diabetes and 23.51 million over the age of 64 have been diagnosed with pre-diabetes [28]. Pre-diabetes cases are especially important to predict future burden and prevalence of diabetes.

The American Diabetes Association reports that the total cost for diabetes and prediabetes in the United States in 2015 was \$322 billion. Additionally, the average economic burden of medical care for individuals with diabetes is 2.3 times higher¹ compared to individuals without diabetes. Health care cost of diabetes care continues to increase; from 2003 to 2013 the price of insulin increased 3-fold thus demonstrating the growing economic burden of diabetes [6,7,28]. The cost of care for both individuals with diabetes and for the nation are points of concern especially for those affected by poverty, lack of adequate access to health care, and minority populations.

The increasing burden of diabetes creates an urgency in understanding the components of its original signal throughout time [6,7,27-29]. Diabetes mortality requires non-parametric analysis to accurately capture its long-term and seasonal fluctuations. Classic time series analysis is not able to handle signals under non-stationary and noisy conditions, however previous studies utilize parametric time-series analysis, thus revealing a current gap in knowledge of diabetes mortality across time. Furthermore, prevention and control of diabetes currently lies upon drastic lifestyle changes, as seen in current and past prevention trials, such as DPP [23]. Lifestyle changes are especially burdensome for those with lack of access to healthcare and insurance, and among vulnerable populations. Thus, it is essential to identify components of adverse diabetes events that are a result of an individual's environment such as radiation.

Methodology

Data

Diabetes mortality data was extracted from the Center for Disease Control Wide-ranging Online Data for Epidemiologic Research (CDC WONDER) database [30]. Monthly counts of total deaths were filtered by International Classification of Diseases, 10th Revision for diabetes (ICD-10: E10-E14). Diabetes mortality under ICD-10: E10-E14 accounts for all diabetes including Type I and Type II. However, only 5% of all diabetes cases are classified as Type I [22,24]. Monthly counts of diabetes-related deaths were aggregated and compiled per year from January 1999 through December 2015. Once all years were merged together, exploratory analyses were performed utilizing both raw data and log-transformed data.

Both raw and log-transformed time series plots displayed signs of seasonality in addition to a non-linear trend. Due to lack of stationary and linear trends in the data, classic parametric time series analysis tools would not yield robust results in comparison to KZ filters. All analyses were conducted in RStudio.

Solar activity is known to have a strong correlation with irradiation, thus prompting the use of ground-based radiation as a way to quantify the relationship between solar activity, such as sunspots, and diabetes mortality in time [2,25-26]. Ground Total Solar Irradiation (GTSI) data was obtained from the National Solar Radiation Database (NSRDB) [31]. Data availability spanned from 1961-2010 and was collected hourly. Summary data is available which compiles hourly data into monthly averages of ground-based total solar activity measured in W/m2. Monthly summary data was exported from a station located in Albuquerque, New Mexico. A subset from January 1999-December 2010 was taken to directly correspond to the diabetes mortality time series.

KZ-filters

The Kolmogorov-Zurbenko (KZ) filter is a low-pass filter with two parameters: m, the length of the moving average window and k iterations of the desired moving average [1,2]. The KZ filter was utilized rather than a standard Fast Fourier Transform (FFT) due to its ability to perform well for noisy and non-stationary data [1,2,12-15,32-34].

Extensions of the KZ filter used in this analysis include the KZ Fourier Transformation (KZFT) and the KZ Adaptive (KZA) filter. The KZFT filter is used for non-stationary data with evidence of periodicities as a result of seasonality and noise in order to reconstruct periodic signals [1,2]. The KZFT filter uses parameters m, length of moving average, and k iterations at a specific frequency, f, identified by spectral analysis [1,2,35-37].

The KZA filter is used to identify change points or breaks in noisy and non-stationary data. The KZA adapts the filtering window of the original KZ filter. Rather than over-smoothing the area where the break occurs, the KZA filter adapts the filter window to zoom in on the break point allowing for clear interpretation of a break point in the presence of a signal suppression due to noise [1-2,38,39].

Diabetes mortality methods

After aggregating monthly diabetes mortality data from the CDC WONDER database from January 1999-December 2015, nonparametric time series analysis was conducted using the kz package in R [30,40]. Spectral analysis in addition to KZ periodograms were used to confirm seasonality and identify long-term trends. Seasonality, long-term trend, and noise were extracted separately. The long-term trend was identified by taking a simple moving average of 13 months and two iterations.

Seasonality (Se) was isolated and smoothed by the following method:

 $Se = KZ(X_i - KZ(X_i, 13, 2), m = 3, k = 1).$

The remaining noise was plotted using the following:

$$\varepsilon = X_t - KZ(X_t, 13, 2) - Se.$$

¹ This estimate is adjusted for age group and sex.

Identification of a change point in the long-term trend utilized the KZFT function in R at the frequency, $f = \frac{1}{12}$ [41]. The KZA package in R identified a break point in the data to correspond with the KZFT change point [42]. Once a change and break point were detected, monthly means before and after the change point were plotted and analyzed.

Analysis of the residuals included obtaining spectra of the logtransformed and raw residuals and assessing peaks present in the spectra plots. The proportion of noise was quantified by obtaining a ρ^2 value from the linear model of the raw time series data modeled by the seasonal component and long-term component. Subtracting 1 from the ρ^2 associated with this linear model yields the proportion of variation (Var_{noise}) attributed to the residuals.

$$%Var_{noise} = 1 - \left[\rho^{2} \left(Linear Model\left(Y_{t} \sim Se + KZ\left(X_{t}, 13, 2\right)\right)\right)\right]$$

Final plots reflect the results of variation due to noise and smoothed seasonality. Normal QQ-plots of the residuals from the noise component were plotted in R to assess the overall signal consisting of the long-term trend and seasonality with respect to the original signal.

TSI methods

Monthly summary data of average TSI was imported into R. Data availability included 1961-2010, however, to directly align with the diabetes mortality data a subset of January 1999 through December 2010 was used for the cross-correlation analysis. The long-term trend of radiation ($LT_{\rm TSI}$) was isolated by removing annual seasonality and two-year cycles. These short-term 2-4 year cycles are a result of regional fluctuations as a component of El Nino [1,2,25,26]. These fluctuations should be smoothed for the use of comparing TSI to diabetes mortality across the United States because they reflect regional differences within the United States.

Annual seasonality was first removed with a simple moving average of 13 months with 2 iterations. The resulting time series had a two-year short-term trend due to El Nino which was removed with a moving average of 23 months and 2 iterations. This yielded an isolated longterm trend of TSI from 1999-2010:

$$LT_{TSI} = KZ(KZ(TSI_{t}, m = 13, k = 2)) = 23, k = 2).$$

This long-term trend was utilized to obtain $\rho 2$ between TSI and diabetes mortality using the cross-correlation function in R [43].

Cross-correlation of TSI and diabetes mortality

The cross-correlation function quantifies the correlation between two time series signals, in this case, TSI and diabetes mortality. The cross-correlation function accounts for possible lag or latency between the two signals of interest. Data points from the opposing datasets are paired at each time point and counted backward or forward by t steps, or t lag. After time points are paired, cross-correlations are calculated starting at t=0, which implies that no latency or lag exists between the two time series. Cross-correlations may be calculated from t=0 to any large value to account for a long latency period between diabetes mortality and TSI. An appropriate upper limit for t can be obtained with several methods, such as maximization between two peak correlations while also accounting for contextual latency of diabetes mortality. Diabetes-related mortality usually occurs after 15 years of diagnosis, and among an elderly population. In this analysis, lag, t, was estimated for the entire data availability, 144 months or 12 years. Final analysis only includes the cross-correlation plot where ρ^2 is maximized at *t*=0 [25,26].

Long-term trends were isolated for diabetes mortality and TSI to quantify the cross-correlation. These ensured results were not biased by underlying components such as seasonality and the El Niño oscillations present in TSI time series. Solar activity measures, such as radiation, are not log-transformed in practice. Thus, the cross-correlation function utilizes original scales for both diabetes mortality and TSI for consistency.

Results

Diabetes mortality results

Spectral analysis confirmed a seasonal cycle of 12 months and additionally a long-term cycle of 102 months which corresponds to 8.5 years (Figure 1). The strength of the seasonality component can be seen in the power compared to the first prominent spike corresponding to the long-term trend. Small spikes in the spectra are a supplement of seasonality due to its prominence as seen in the power. Peaks in spectra prompted separation of the seasonal and long-term trend.

The first cycle of the long-term component reflects an initial 10% increase in diabetes deaths from December 1999 through January 2005, and a 10% decrease in diabetes deaths from January 2005 until the end of the first 8.5-year cycle, June 2008 (Figure 2a). The seasonal component (Figure 2b) revealed increased frequencies of diabetes mortality during winter months, which corresponds to lowest sun activity. In contrast, decreased diabetes mortality corresponds to warmer months when sun activity is highest. Thus, the relationship between sun activity and diabetes mortality would be expected to have a negative correlation coefficient. Isolation of seasonality and the long-term trend accounted for approximately 85% of the true signal reflected in the maxima of the residuals (Figure 2c).

Due to the large dominant frequency in spectra at 12 months, $f = \frac{1}{12}$, (Figure 1), a KZFT filter was used to analyze the log-transformed data at $f = \frac{1}{12}$. The KZFT filter displayed evidence of a phase shift at April 2007 which corresponds to a time-point index 100 out of total 204 time points (Figure 3a). A break point was identified to occur at December 2008 using the KZA filter, occurring directly after the first long-term cycle of 8.5 years (Figure 3b).



Figure 1: Depicts the raw spectra of diabetes mortality data with spikes highlighted in red at 0.0093 ($f = \frac{1}{102}$) and 0.083 ($f = \frac{1}{12}$), thus indicating long-term trend of 8.5 years and seasonality.



Figure 2: Depicts the log-transformed raw time series of diabetes mortality with a moving average of 13 months iterated 2 times using KZ(13,2) (a), the logtransformed seasonal component time series after removing the long-term trend, KZ(13,2), and smoothing for noise, KZ(3,2) (b), and the remaining noise after smoothing and removal of log-transformed long-term trend and log-transformed seasonality components (c).



Figure 3: Highlights the change point at time point 100, which corresponds to the year April 2007, after analyzing the KZFT(m=204, k=1, f=0.083) for the logtransformed diabetes mortality data (a), and the KZ decomposition of the log-transformed time series. The KZA filter, bottom panel, reflects the change point starting just before 2009 (b).

With the identification of the change point at December 2008, which aligned with the 10% increase after December 2008 in the long-term component, aggregated monthly counts of diabetes mortality were plotted separately (Figure 4). Aggregated monthly counts of diabetes death before and after December 2008 show increased amounts of diabetes deaths after the change point, especially during warmer months and January (Figure 4).

Figure 2c displayed some residual seasonality within the noise component, thus prompting further investigation. Spectra were plotted and analyzed for the residuals. Pronounced peaks are revealed at 2.5 and 3 months corresponding to, $f = \frac{1}{3}$ and $\frac{1}{2.5}$, respectively. These frequencies correspond to the small fluctuations seen in the unsmoothed seasonal trend, which are a result of the annual seasonal trend. Small fluctuations are to be expected due to the harmonic sine wave result

of KZFT at $f = \frac{1}{12}$. Diabetes mortality resembles a sin wave but does not fit perfectly in nature, thus leaving minimal seasonality leakage in the residuals. That leakage is very small as can be noticed from spectra in Figure 1, Normal QQ plot in Figure 5 proves that residuals resemble high quality noise with the corresponding correlation coefficient of 0.965. Thus, the residuals contribute approximately ±5% toward the total signal.

Noise assessment was done using a Normal Quantile-Quantile (QQ) plot of the residuals from the noise component of the log-transformed signal, thus the plot reflects the removal of annual seasonality and sun activity, 8-11 years, trend. The Normal QQ plot of the residuals shows a good fit to a normal distribution and only contains high frequencies (Figure 5).







Figure 5: Normal QQ-plot of the residuals from the noise component after removing long-term trend, oscillation of 8-11 years periodicity, and annual seasonality from original data on the log-scale. Log scale provides differences in approximate percentage scale; thus the range of residuals contribute approximately 5% toward the total signal.

Analysis of the residuals for both the log-transformed and raw data reveal that the log-transformed data stabilize the variance and have an improved signal-to-noise ratio. The log-transformed signal-to-noise ratio was 5.67 compared to 5.09 for the raw data.

Cross-correlation results

The spectral peak corresponding to an 8.5-year long-term trend reveals a relationship with solar activity. The unique signal of solar activity has been proven previously which corresponds to an 8-11 years cycle [1,2,25,26]. Similar spectra peaks in solar activity and diabetes mortality prompted further investigation quantifying the correlation between the two signals. TSI was used as the measure of solar activity, which is proven to be strongly correlated with sun activity [2,25,26]. The long-term trend plot of diabetes mortality overlaid with the longterm trend of TSI indicate a strong inverse relationship (Figure 6a). During 2004-2006, diabetes mortality is at its peak whereas TSI is at its lowest. The cross-correlation coefficient shown in Figure 6b quantifies the inverse relationship in Figure 6a. Cross-correlation results found

J Biom Biostat, an open access journal ISSN: 2155-6180

that the correlation coefficient is maximized at t=0, with $\rho^2=-0.76$ (Figure 6b). A high negative correlation coefficient confirms a strong inverse relationship of diabetes mortality and TSI. Thus, when TSI decreases, diabetes mortality will increase (Figures 4 and 6).

Conclusion

Spectral analysis reflects a long-term trend and seasonality of mortality among diabetes deaths with prominent peaks at frequencies corresponding to 8.5 years and 1 year. Small peaks occurring after 12 months in the spectra of the log-transformed data are a result of the strong presence of seasonality. It has been previously proven that sunspot activity has a unique frequency peak in spectra which corresponds to the spectral peak at 8.5-years in diabetes mortality [1,2]. This proof of solar activity prompted further investigation between diabetes mortality long-term trend and solar activity long-term trend [2].

The corresponding 8.5-year cycle depicts the first cycle from January 1999 through June 2008 and the second cycle starting after June 2008 through the data cut-off, December 2015. The first cycle of the long-term component reflects an initial 10% increase in diabetes deaths from December 1999 through January 2005, and a 10% decrease in diabetes deaths from January 2005 until the end of the first 8.5-year cycle, June 2008. The second cycle, which starts after June 2008, shows approximately a 15% increase in diabetes deaths from January 2015 and decreases until the data-cutoff in December 2015.





The difference from the first and second cycles of the long-term trend reflect an especially sharp increase in deaths around January 2015 in comparison to the peak in the first 8.5-year cycle. This increase in diabetes deaths could be the result of reporting changes in ICD-codes and a reflection of a very latent disease that progresses with age. The complexity of diabetes and its many co-morbidities make it hard to classify as the underlying cause of death. The American Diabetes Association reports that diabetes is likely under classified as the cause of death due to the results of studies which state that between 35-40% of the deceased with diagnosed diabetes had diabetes listed on their death certificate, however, only 10-15% of these death certificates listed diabetes as the underlying cause of death.

The spectra show the most prominent peak at the corresponding 1-year frequency. The magnitude of the sharp peak indicates the strong presence of seasonality among diabetes mortality in the United States. The seasonality component reveals increased diabetes mortality during winter months and the lowest incidence of diabetes mortality during June through September months. These results correspond to previous studies that previously address seasonal differences in diabetes mortality in which diabetes mortality is increased in during winter in comparison to summer months [9-11].

Seasonal differences in diabetes mortality correspond to seasonal trends in sun activity. June is known to have the highest sun activity and lowest sun activity occurs in December or January. Figure 4 shows highest diabetes mortality occurring during winter months, December and January and lowest diabetes mortality during summer months. Seasonal sun activity trends are strongly correlated with seasonal radiation trends, thus TSI was used to quantify the negative correlation between solar activity and diabetes mortality [1,2,25,26].

Some studies hypothesize that diabetes mortality seasonal differences are attributed to varying weather patterns as well as latent adverse effects of poorly managed glycemic levels. Tseng et al. found seasonal variations in hemoglobin A1c levels, where high HbA1c levels were found to occur during winter months in comparison to summer months [9]. HbA1c levels have been shown to be correlated with negative adverse effects for diabetes mellitus. Additionally, long-term uncontrolled HbA1c levels often lead to co-morbidities closely aligned with diabetes such as chronic kidney disease and cardiovascular disease [9,16-24,28].

After isolating seasonal and long-term trends, the residuals reflected that 85% of the true signal of the log-transformed diabetes mortality time series data was covered. However, there still appeared to be slight seasonality present within the residuals (Figure 2c). Thus, $1-\rho^2$ value was calculated to quantify the proportion of variation of the true signal that was attributed to the residuals. The $1-\rho^2$ was 15% for the log-transformed data, thus, the residuals attributed 15% of the variation not accounted for by the seasonal and long-term components. Supplementary peaks of seasonality occurring at 2.5-3 months reflect the strength of seasonality presence among the residuals.

A sharp increase in diabetes mortality after 2008 led to use of the KZFT and KZA filter to confirm a change point in the sinusoidal signal of diabetes mortality. The KZFT filter identified a change in frequency of the sin wave for log-transformed diabetes mortality time series data in 2008. The KZA filter identified a change or break-point in noisy and nonstationary data starting in 2008. This break point corresponds to the phase shift found in the KZFT filter as well as the sharp increase after 2008 in the long-term trend plot. The phase shift and break point may correspond with changes in reporting for diabetes as a cause of

death on death certificates or a change in the way ICD codes were recorded. In 2008, there were new revisions made for diabetes mellitus with additional categories and sub-categories allowing additional classification under diabetes mellitus as a cause of death. However, as of October 2015, ICD-10 codes are now used in place of ICD-9 codes due to the ICD-10 codes including twice the categories and classifications in comparison to ICD-9 codes [44].

Aggregate monthly means taken before and after 2008, where the phase shift and change point approximately occurred, show an overall increase across all months. However, January and summer months, May through September, had a much higher increase in diabetes mortality after December 2008. This reflects the phase shift and change point in diabetes mortality and could be the result of ICD code changes as well as an increasing elderly population due to the "Baby Boomer" generation [24,45]. The increase in diabetes diagnoses among all populations, especially in the young population, seemingly may influence the phase shift and sharp increase in death as well as administration or policy changes. In addition to a limited time frame of data points, the reports of diabetes as an underlying cause of death may be underreported due to diabetes' close relationship with cardiovascular events.

The spectrum peak associated with an 8.5-year cycle of diabetes mortality prompted further investigation of the relationship between solar activity and diabetes mortality. Furthermore, annual seasonality and aggregate monthly means plots reflect increased diabetes mortality during winter months when solar activity is lowest and decreased diabetes mortality during summer months when solar activity is highest. This a priori knowledge of solar activity in conjunction with findings from the KZ-filter analysis on diabetes mortality led to the quantification of the negative relationship between the two signals. Cross-correlation between diabetes mortality and TSI resulted in time lag equal to 0 with a -0.76 correlation coefficient. The strong negative correlation is the first quantified measure of solar activity's relationship to diabetes mortality. More investigation regarding this negative feedback relationship between TSI and diabetes mortality across time is needed to solidify this finding. However, the strength in correlation indicates the real presence of a relationship between the two time series. Solar activity trends in time may indicate other seasonal environmental changes that directly stimulate adverse events among those with diabetes, eventually leading to increased diabetes mortality during months following exposure to high solar activity.

In conclusion, the discovery of annual seasonality indicates that future studies should control for diabetes mortality differences during summer and winter months. Additionally, future studies should confirm and control for seasonality in HbA1c or additional measures that indicate poorly maintained diabetes. Prolonged mis-managed HbA1c is associated with higher probability of adverse events among diabetics [9,24]. Thus, controlling for seasonality among HbA1c measures and duration of diabetes may lead to less biased outcomes.

The discovery of corresponding spectra peaks in diabetes mortality and solar activity prompted the first investigation of their relationship. This was the first study, to our knowledge, that has quantified the relationship between solar activity and diabetes mortality. The strong negative correlation discovered has strong implications of the environmental role in diabetes mortality. Using solar activity makes this correlation coefficient less likely to be an effect modifier due to the role of solar activity in regulating conditions such as temperature fluctuations and seasonality. Future studies may find this useful to control for environmental conditions related to diabetes adverse events, such as mortality.

Page 6 of 8

Limitations

Limitations of this analysis include the lack of data prior to 1999 and after December 2015. The long-term trend could have been more insightful with a longer time-frame of data, especially with a phase shift of the sinusoidal wave occurring halfway through the data. The data used to compile all diabetes mortality counts within the United States consisted of diabetes mortality with the underlying cause of death as diabetes mellitus, which could refer to both Type 1 and Type 2 diabetes mellitus. However, over 90-95% of all diabetes cases in the United States are type 2 diabetes mellitus, thus these rarer cases would not have much influence on the results of this analysis. The complexity of diabetes as a disease with its multiple co-morbidities especially those with advanced diabetes create a challenge for citing diabetes as the cause of death on death certificates.

Papers that cite hemoglobin A1c levels fluctuate in the same cycle as diabetes mortality, high in the winter and lower during spring and summer, correspond to decreased physical activity during winter months [9]. This analysis could benefit from utilizing spatial analysis to assess differences between states with cold and hot seasons compared to states with less seasonal variation. Hemoglobin A1c levels, especially uncontrolled, have a significant effect on health outcomes for those with diabetes. Individuals diagnosed with diabetes for longer amounts of time and who have gone through multiple winter seasons with uncontrolled glycemic levels may lead to the increase in diabetes mortality during winter months.

To our knowledge, no other studies have published findings utilizing KZ filters and investigating the relationship between solar activity and diabetes mortality. Additionally, there was a lack of timeseries analysis using other classic methods assessing diabetes mortality thus making it difficult to compare and validate my results with others' research. Most research done with diabetes and time series analysis were outdated, utilized parametric methods, were done in countries outside of the United States, or had focused population groups that could not generalize beyond their target population. Cross-correlation results only reflect one-cycle from the long-term components of solar activity and diabetes mortality due to lack of data availability. Further investigation is needed with more time points that capture more than one cycle to assert more robust inference from this quantified relationship.

References

- 1. Yang W, Zurbenko I (2010) Kolmogorov-Zurbenko Filters. Wiley Interdisciplinary Reviews: Computational Statistics.
- Zurbenko IG, Smith D (2017) Kolmogorov-Zurbenko filters in spatiotemporal analysis. Wiley Periodicals, Inc. WIREs Computational Statistics.
- Zurbenko IG (1986) The spectral analysis of time series. In North-Holland series in statistics and probability. Elsevier Science Pub. Co, Amsterdam, New York.
- 4. Murray R (1985) Stationary Sequences and Random Fields. Springer, Boston.
- 5. Nason GP (2008) Wavelet Methods in Statistics Using R. Springer, New York.
- Yang W, Dall TM, Halder P, Gallo P, Kowal SL, et al. (2013) Economic Costs of Diabetes in the U.S. in 2012. Diabetes Care 36: 1033-1046.
- Dall TM, Zhang Y, Chen YJ, Quick WW, Yang WG, et al. (2010) The Economic Burden of Diabetes. Health Affairs (Project Hope) 29: 297-303.
- Yang J, Yin P, Zhou M, Ou CQ, Li M, et al. (2016) The Effect of Ambient Temperature on Diabetes Mortality in China: A Multi-City Time Series Study. Science of the Total Environment 543: 75-82.
- Tseng CL, Brimacombe M, Xie M, Rajan M, Wang H, et al. (2005) Seasonal Patterns in Monthly Hemoglobin A1c Values. American Journal of Epidemiology 161: 565-574.

 Ostro B, Broadwin R, Green S, Feng WY, Lipsett M (2006) Fine Particulate Air Pollution and Mortality in Nine California Counties: Results from CALFINE. Environmental Health Perspectives 114: 29-33.

Page 7 of 8

- Sun S, Qiu H, Ho KF, Tian L (2016) Chemical Components of Respirable Particulate Matter Associated with Emergency Hospital Admissions for Type 2 Diabetes Mellitus in Hong Kong. Environment International 97: 93-99.
- Shumway RH, Stoffer DS (2006) Time Series Analysis and Its Applications with R Examples. Design.
- Brillinger DR (1981) Time Series: Data Analysis and Theory. Expanded Edition. San Francisco.
- Zurbenko IG, Sowizral M (1999) Resolution of the Destructive Effect of Noise on Linear Regression of Two Time Series. Far East Journal of Theoretical Statistics 3: 139-157.
- http://libproxy.albany.edu/login?url=http://search.ebscohost.com/login.aspx?d irect=true&db=msn&AN=MR1704700&site=eds-live&scope=site
- Tsakiri KG, Zurbenko IG (2010) Determining the Main Atmospheric Factor on Ozone Concentrations. Meteorology and Atmospheric Physics 109: 129-137.
- Clarke PM, Gray AM, Briggs A, Farmer AJ, Fenn P, et al. (2004) A Model to Estimate the Lifetime Health Outcomes of Patients with Type 2 Diabetes: The United Kingdom Prospective Diabetes Study (UKPDS) Outcomes Model (UKPDS No. 68). Diabetologia 47:1747-1759.
- Hayes AJ, Leal J, Gray AM, Holman RR, Clarke PM (2013) UKPDS Outcomes Model 2: A New Version of a Model to Simulate Lifetime Health Outcomes of Patients with Type 2 Diabetes Mellitus Using Data from the 30 Year United Kingdom Prospective Diabetes Study: UKPDS 82. Diabetologia.
- Bebu I, Braffett BH, Busui RP, Orchard TJ, Nathan DM, et al. (2017) The Relationship of Blood Glucose with Cardiovascular Disease Is Mediated over Time by Traditional Risk Factors in Type 1 Diabetes: The DCCT/EDIC Study. Diabetologia.
- Mohammedi K, Woodward M, Hirakawa Y, Zoungas S, Colagiuri S, et al. (2016) Presentations of Major Peripheral Arterial Disease and Risk of Major Outcomes in Patients with Type 2 Diabetes: Results from the ADVANCE-ON Study. Cardiovascular Diabetology.
- O'Connor PJ, Narayan KM, Anderson R, Feeney P, Fine L, et al. (2012) Effect of Intensive versus Standard Blood Pressure Control on Depression and Health-Related Quality of Life in Type 2 Diabetes: The ACCORD Trial. Diabetes Care.
- Margolis KL, O'Connor PJ, Morgan TM, Buse JB, Cohen RM, et al. (2014) Outcomes of Combined Cardiovascular Risk Factor Management Strategies in Type 2 Diabetes: The Accord Randomized Trial. Diabetes Care.
- The ADVANCE Collaborative Group (2008) Intensive Blood Glucose Control and Vascular Outcomes in Patients with Type 2 Diabetes. (ADVANCE Trial). The New England Journal of Medicine.
- Diabetes Prevention Program Research Group (2015) Long-Term Effects of Lifestyle Intervention or Metformin on Diabetes Development and Microvascular Complications over 15-Year Follow-up: The Diabetes Prevention Program Outcomes Study. The Lancet. Diabetes & Endocrinology.
- Zinman B, Lachin JM, Inzucchi SE (2015) Empagliflozin, Cardiovascular Outcomes, and Mortality in Type 2 Diabetes. New Engl J Medicine.
- Valachovic E, Zurbenko I (2014) Skin Cancer, Irradiation, and Sunspots: The Solar Cycle Effect. BioMed Research International.
- Valachovic E, Zurbenko I (2017) Multivariate Analysis of Spatial-temporal Scales in Melanoma Prevalence. Cancer Causes and Control.
- 28. American Diabetes Association (2017) Diabetes Statistics.
- American Diabetes Association (2017) Standards of Medical Care in Diabetes -2017 Supplement 1. Diabetes Care S88-S98.
- National Diabetes Statistics Report (2017) Estimates of Diabetes and Its Burden in the United States. Public-use data file and documentation.
- 31. Centers for Disease Control and Prevention, National Center for Health Statistics. Multiple Cause of Death 1999-2015 on CDC WONDER Online Database, released December, 2016. Data are from the Multiple Cause of Death Files, 1999-2015, as compiled from data provided by the 57 vital statistics jurisdictions through the Vital Statistics Cooperative Program.

- Wilcox S (2010) National solar radiation database 1991-2010 update: user's manual. NREL/TP-581- 41364.
- Yang W, Zurbenko I (2010) Nonstationarity. Wiley Interdisciplinary Reviews: Computational Statistics.
- Wise EK, Comrie AC (2005) Extending the Kolmogorov-Zurbenko Filter: Application to Ozone, Particulate Matter, and Meteorological Trends. Journal of the Air and Waste Management Association.
- Zurbenko IG, Rao ST, Henry RF (1995) Mapping ozone in the Eastern United States. Environ Man 24-30.
- Zurbenko IG, Potrzeba-Macrina AL (2013) Periods of Excess Energy in Extreme Weather Events. Journal of Climatology.
- Neagu R, Zurbenko I (2003) Algorithm for Adaptively Smoothing the Log-Periodogram. Journal of the Franklin Institute 340: 103-123.
- Zurbenko IG, Potrzeba-Macrina AL (2010) Tidal Waves in the Atmosphere and Their Effects. Acta Geophysica 58: 356-373.

- Zurbenko I, Porter PS, Rao ST, Ku JY, Gui R, et al. (1996) Detecting Discontinuities in Time Series of Upper-Air Data: Development and Demonstration of an Adaptive Filter Technique. Journal of Climate 9 (12 II): 3548-3560.
- Close B, Zurbenko IG (2008) Time Series Analysis by KZA of the Fatal Analysis Reporting System. Proceedings of the Joint Statistical Meeting. Denver, Colorado.
- 41. Close B, Zurbenko IG (2004) kz: Kolmogorov-Zurbenko Filter. R package.
- 42. Yang W, Zurbenko IG (2006) kza: Kolmogorov-Zurbenko Fourier Transform and Application. R package.
- Close B, Zurbenko IG (2004) kza: Kolmogorov-Zurbenko adaptive algorithm for the image detection. R package.
- 44. Stoffer D (2016) Astsa. R package.
- 45. National Center for Health Statistics (2017) ICD-9-CM Tabular Addenda.

Page 8 of 8