Investigation of the Maximum Association for Suicide Rate and Social Factors Using Computer Algorithm

Fukuda, K. [1] and T. Takaoka [2]

[1] Environmental Science Programme, Department of Mathematics and Statistics, and Computer Science & Software Engineering, University of Canterbury, Christchurch
[2] Department of Computer Science & Software Engineering, University of Canterbury
Email: k.fukuda@math.canterbury.ac.nz

Keywords: *Algorithm, Factor, Maximum subarray, Socioeconomics, Suicide*

EXTENDED ABSTRACT

Suicide is a worldwide concern; reducing the suicide rate is a primary role of public, government and communities. A computer algorithm, the *K*-Maximum Subarray (*K*-MSA), is applied for the first time to investigate the maximum relationship between the total New Zealand population of suicide deaths (1983-2002) by age group, with four selected social factors; number of bankruptcies, unemployed people, divorces, and orphans and unsupported child beneficiaries, respectively. The range of ages from 5 to 80 years is divided into five-year groups, plus a "more than 85 years old" group. Social factors are classified into four classes (*low*, *med*, *high* and *very high*) based on their quartile ranges.

The advantage of using the *K*-MSA algorithm is that it detects threshold values (which age groups associate most with social factors) by maximising the sum of the elements of selected subarrays of a two-dimensional array; this region is called a maximum subarray. The *K*-MSA has two parameters, *K* and *w*. The *K*-MSA detects *K* maximum subarrays. The weight parameter, *w*, is generally fixed as the average of the values in the full array, but a new approach is introduced to allow extracting smaller subsets within the large maximum subarray by changing *w* values (using percentiles from 50, 75, 85 and 95). This results in providing the detailed association of the specific suicide age groups and social factors.

The *K*-MSA with lower *w* values (e.g., *w*=50 and 75 percentile) successfully identifies the general trends. From the largest maximum subarray (at *k*=1), a similar threshold pattern of social factors between bankruptcies and unemployment rate, and divorces and orphans and unsupported child beneficiary numbers was found respectively for females and males. This may suggest that these two similar factors may impact the suicide rate in a similar manner, but different sexes respond differently. For the age, the maximum subarray applied to all four factors, shows two female age groups; 15-54 and 55-59 years old (plus a third, with unemployment rate: 60-64 years old), and one male age group, 15-59 years old, suggesting that the suicide rate is slightly extended out for females than males with the studied social factors.

Further changing the weight parameter (*w*=85 and 95) successfully detected detailed association of the suicide age and social factor level. Overall, very high numbers of divorces are found to be associated most with the high suicide rate among other factors. For female, the extreme divorce rate associates with the wider female age group, 15-44, especially for younger females, 15-24 years old. Also, this later age group and the late 30s associates with high orphan and unsupervised beneficiary numbers, late 30s to early 40s and teenager with extremely high bankruptcies, and the early 40s and late 20s with high unemployed rate. For male, very high divorce number associates firstly with 30s and then 20s, med to high orphan number and med to very high unemployed rate with 20s to perhaps up to 34 years old, and the late 30s to early 40s, and 20s associate with extremely high and med to high bankruptcies rate, respectively.

An experimental investigation using the *K*-MSA and a new approach, changing the weight parameter successfully provide detailed associations for the specific suicide age groups and social factor levels. Results do not identify the direct causal factors for the individual suicidal behaviour, but represent the national trends. The *K*-MSA method can be further evaluated with different factors as a data mining tool to help minimising the future suicide rate by setting up a specific policy for the specific age groups for different sexes, and perhaps the socioeconomic status as background information.

1. INTRODUCTION

Suicide is a worldwide concern, and reducing the suicide rate is a primary role of public, government and communities. However, identifying and quantifying causal factors cannot be a simple task as many hidden factors that can be involved, and potential causal factors may be indefinable (e.g., genetic, family, health, personal and economic issues). The Ministry of Health (MOH, 2006) reports that the New Zealand suicide rates in 2003 for all age groups for males and females are respectively 6th and 4th highest among OECD countries, following Finland, Japan and France. Historically, the suicide rate in New Zealand increased from the mid-1980s (12.0 deaths per 100,000 population or an average of 360 deaths per year) to the mid-1990s (16.7 deaths per 100,000 population or an average of 560 deaths per year), and decreased since then. Generally, a higher suicide rate is found for males than females between 1983 and 2003, with a peak in suicide deaths from 1990-1992 (4.2 male deaths for every female death), but it has decreased to a ratio of 3.2 male deaths to every female death since then.

The public and policy have tended to focus on preventing suicide for school-aged young people (Beautrais, 2003b, MOH, 2006) due to the historical trend of high suicide rate among 15-24 years old. However, Beautrais (2003a,b) suggests a need to shift/adjust the policy to pay greater attention to the 20-29 age range and the older age range, 20-45 years, because only approximately 15% of youth suicides are from school-aged young people – two thirds of youth deaths, in fact, occur in the 19-24 age range. Also, the age group of 20-45 is found as the highest risk for suicide for both males and females (e.g., 75.6% of the 516 suicide deaths in 1999 are adults aged over 25).

Beautrais (2003b) also summarised that suicidal behaviour is closely associated with six conceptually related factors; genetic and biologic factors, social and demographic background, childhood adversity, personal characteristics, life stress and mental health factors and these associate differently among various age groups. Young people suicide rate increases among those who are with poor educational qualifications and family background with low socioeconomic status, and a link is found between socioeconomic status and mental health and adjustments in adolescence. Besides this, older adults who have never married, are divorced or widowed, and live alone, have more likelihood of suicidal behaviour (see details in Beautrais, 2003b).

The accurate suicidal behaviour can be assessed with individual background of potential factors, though it may require significant effort to collect information and investigate confounding factors (e.g., personal issues, childhood adversity, education and family background, and mental status). Generally, reports for the suicidal behaviour are assessed by a time series analysis (e.g., linear regression and moving average; MOH 2006). However, this experimental study is aimed to examine the general idea of the association between national scale suicidal behaviour and selected socioeconomic factors using the national annual New Zealand suicide rate and New Zealand census-linked data using a computer algorithm for the first time. Machine learning techniques includes data mining, have been discovered as powerful tools to investigate the new filed of researches such as air pollution, climate and health (Fukuda and Takaoka, 2007), since the computer algorithm can extract knowledge about the data set efficiently.

The advantage of using the *K*-MSA algorithm is that it detects threshold values (which age groups associate most with social factors) by maximising the sum of the elements of selected subarrays of a two-dimensional array. Also, the *K*-MSA, does not require assumptions or significant data pre-processing to be comparable. Thus, without directly defining the causes of suicide, *K*-MSA successfully detects the maximum association of suicidal rate and social factor by incorporating the age of suicide for males and females respectively. In this study, the new approach is introduced to allow detecting a detailed maximum subarray by changing the weight parameter (w). This consequently successfully provides detailed information on specific suicide age ranges and social factors relationship.

2. METHODS

Bae and Takaoka (2005) developed the *K*-MSA algorithm with a slight modification from Kadane's algorithm that only detects the maximum subarray in a one-dimensional array. The *K*-MSA detects *K* maximum subarrays in the given two-dimensional array, constructed by the horizontal index x of the given array, corresponding to the selected social factor, which is described by four classes (low, med, high and very high) and the vertical index y for the suicide age groups, shown in Figure 2. This two-dimensional array takes the third data point z (suicide rate) as input. Each value of z is linked by the same time period of occurrence of the suicide and the change of the social factors. Results provide knowledge on how the maximum suicide rate that occurred within the

certain age group of the particular year associates with the socioeconomic status of that year.

2.1. *K*-Maximum Subarray

Firstly, Kadane's algorithm is explained as follows. Let s be the sum of a candidate maximum subarray $a[k..l]$ and t be the tentative sum of a subarray $a[j..\ i]$, while the specified w value, defined by the weight parameter, is subtracted from each array element (i.e., any real numbers, including negative values). Figure 2 shows how the algorithm scans through the given one-dimensional array by accumulating a tentative sum in t, replacing s by t and updating the position (k, l), when $t > s$ is detected (see also in Fukuda and Takaoka, 2007). When $t<0$ is detected, the accumulation is reset to zero. This algorithm takes $O(n)$ time and gives a maximum subarray $a[k..l]$ of $a[1..n]$,

(k, l) := (0,0); s := -∞; t := 0; j := 1;
for i := 1 **to** n **do begin**
    t := t +$a[i]$;
    if $t > s$ **then begin** (k, l) := (j, i); s := t **end**;
    if $t < 0$ **then begin** t := 0; j := i +1 **end**
end

For the *K*-MSA process, Kadane's algorithm is extended to two dimensions. Let a two-dimensional array $a[1..m,\ 1..n]$ be input data, where the value of each element $a[i, j]$ is similar to the above. The general idea of the two-dimensional maximum subarray is to maximise the sum of the array portion $a[k\ ..\ i,\ l\ ..\ j]$, where (k, l) and (i, j) are index pairs corresponding to the upper-left corner, and the bottom right corner of the subarray. Note that upper case K and lower case k are used for different purposes.

1. For each row k of array a (k≥1)
2. For each row $i \geq k$ of array a
3. Solve the one-dimensional maximum subarray for the strip portion from row k to row i
4. Let the solution be $a[k\ ..\ i,\ \ l\ ..\ j]$
5. Take the maximum of the $m(m$-1)/2 solutions.

Line 3 takes $O(n)$ time by itself, and it is placed in the doubly nested loop by k and i, resulting in $O(m^2n)$ time for line 3 in total. On the other hand, line 5 takes $O(m^2)$ time. When m=n, the total time is $O(n^3)$. This algorithm is extended to an algorithm for the *K*-MSA subarray problem in the following way. Once the maximum subarray is detected, each array element in it is replaced by minus infinity, -∞, (-X, where X is a large number in practice) to obtain the second maximum subarray problem from the modified array and so on. This results in $O(Km^2n)$ time or $O(Kn^3)$ time for m=n. This algorithm and further speed-up is described in Bae and Takaoka (2005).

K-parameter setting

For the *K*-MSA, the user can specify K≥1 to obtain k–th maximum subarray for k=1, …, K. The first maximum subarray (at k=1) has the largest sum value, s, in the two-dimensional array, indicates that the most significant association for the specific age groups of suicide at the selected social factor. Each maximum subarray at k is detected uniquely and does not overlap with other subarray detected at different k, called a disjoint *K*-MSA. This means that once the association is detected between the certain suicide age group and a social factor level (threshold), *this* part is not detected again.

Weight parameter setting

A new approach, the weight (w) parameter setting, is applied for the first time to capture different aspects of the large maximum subarrays by providing more detailed maximum subarrays with further w values (see Figure 2). Let the horizontal co-ordinate x be the social factor level and the vertical co-ordinate y represent the suicide age group. Let $Sx[1, .., NSx]$ be a series of social factor levels and $Sy[1, .., NSy]$ be a series of suicide age groups. While the suicide age groups are described

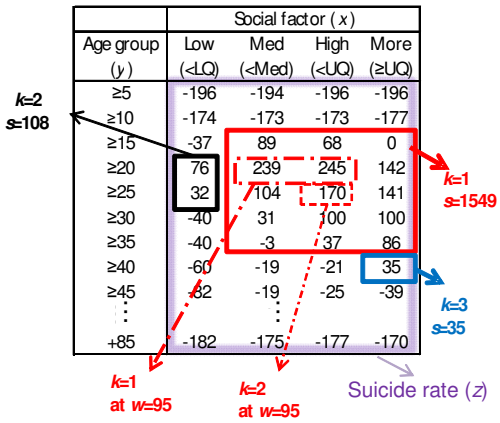


Figure 2. Example of the *K*-MSA matrix from the male suicide rate with bankruptcies rate at w=75. Solid and dotted lines indicate the position of maximum subarrays with w=75 and w=95 respectively. Note that the s value for w=95 is not represented as the matrix is obtained for only w=75.

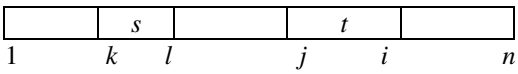


Figure 3. Kadane's algorithm.

for every five years from ≥5 to ≥80 and 85 plus, then Sk = (5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, more) and SBx=17. Practically, when y=5, the age group is from 25 to 29. Let [1, SBx] be the range in which x lies and [1, NSy] be the range in which y lies. Then, a matrix M (of dimension NSx, NSy) is calculated, where $M(i, j)$ is the number of observations that lie in the range of (i, j). Note that the rows i of M correspond to age groups and the columns j correspond to social factor.

The weight parameter, w, is defined by the user's specific P percentile. The P percentile of all the values in Z (suicide rate) is then calculated as,

$$p = \frac{1}{NSx \cdot NSy} \sum_{i}^{NSx} \sum_{j}^{NSy} Z(i, j) \ . \qquad (1)$$

A new matrix M' of the same dimensions as M

$$M'(i, j) = M(i, j) - p \ , \qquad (2)$$

is created as the two-dimensional maximum subarray of M' for the input data array a for the algorithm, described in the previous section.

3. DATA SETS

Twenty years of annual New Zealand national suicide rate data (1983-2003) was taken from Ministry of Health report (detailed in MOH, 2006). The same period of four social factors, consolidated bankruptcies, consolidate (actual) registered unemployment rate, numbers of orphans and unsupervised beneficiaries, and numbers of divorce, are taken from New Zealand Statistics database (detailed description for each factor is in NZ Stats, 2006). Note that each factor is denoted as following in this study: bankruptcy rate, unemployment rate, number of orphans, and number of divorces. These four social factors are selected, because they are fully completed to match the period of the suicide rate time series and are assumed to show some potential associations with suicidal behaviours.

The suicide data were analysed for females (n=2091) and males (n=7353), applying the K-MSA for a single factor with four w values (50, 75, 85 and 95) up to =3, respectively. All input suicide data sets were normalised as to allow comparable results between sexes (normalised total number for female is 8044 and for male is 9427). The age groups were divided into every five years, 17 groups from ≥5, ..., ≥80 to 85+ years old. Each social factor is categorised into four classes based on the quartile range; *low* (< lower quartile), *medium* (<median), *high* (<upper quartiles), and *very high* (≥*high*). The suicide and social factor data sets were constructed by linking each suicide rate to each social factor with the same year. This keeps the annual time trend dependency.

4. RESULTS AND DISCUSSIONS

In Figure 4, detected maximum subarrays are described by a tree structure, which demonstrates how the large maximum subarrays at lower w values are expanded to provide smaller and more detailed maximum subarrays at higher w values. Each tree shows only major findings based on s values above 0 and up to k=3 with w=50, 75, 85 and 95. Interpretations in this section from these major findings describe the associations of more than *high* social factor levels with female and male suicide rate (the *low* and *med* social factor levels are shaded in Figure 4).

4.1. General trends

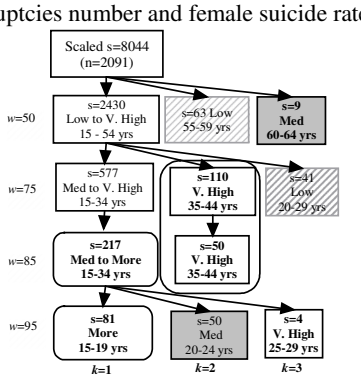
The lower w values (w=50 and 75) identify the general trends. For the factors, the threshold pattern detected by the largest maximum subarray (at k=1), is found to be similar between bankruptcies and unemployment rate, and divorces, orphans and unsupported child beneficiary numbers for each sex, respectively. For example, both of the first maximum subarrays of female bankruptcies (Figure 4-A) and unemployment (Figure 4-C) at w=50 detect the same threshold: age 15-54 and factor more than *low*. The same structure is found at w=75 for these two factors. These similarities over two factors may suggest that the impact on the suicide rate is similar manner, but different sex responses differently. However, note that male divorces and male orphans have slightly different structures at k=1 for w=50 and w=70.

For the age, the maximum suicide age groups are detected from all factors as follows. For females, two age groups, 15-54 and 55-59 years old (but the unemployment rate has a third age group; 60-64 years old) are detected. For male, a single age group, 15-59 years old, is detected, in Figure 4. It suggests that the suicide rate range is slightly extended out for females than males, when these factors are considered.

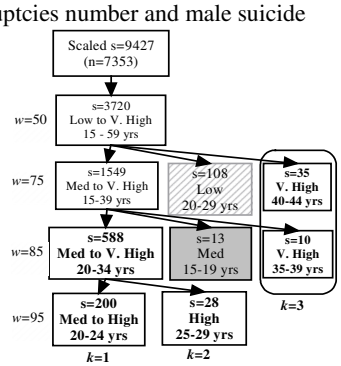
Excluding the broad range of social factor levels (e.g., *low* to *very high*), all maximum subarrays up to k=3 for males and divorce rate (white cells in Figure 4-F) show the association of the higher divorce rate and male suicide rate. It suggests that the high male suicide rate may associate with the high divorce rate and its association is at least time

sensitive (since the *K*-MSA input kept the same year of the suicide and divorce rate). Note that this observation is also same for males and the other social factors. For female, the orphan number (Figure 4-G) shows that half of the maximum subarrays are detected to associate with either *higher* orphan number (=1 and a high *s* value) or *lower* orphan number (*k*=2 and 3 and lower *s* values). The former suggests that the high female suicide rate may associate with the extremely high levels of orphan number. The latter may suggest that the indirect association of orphan number and female suicide rate, though some other factors that perhaps cause the orphan numbers to be *low* or *med* may have a direct association, but it requires further investigation.

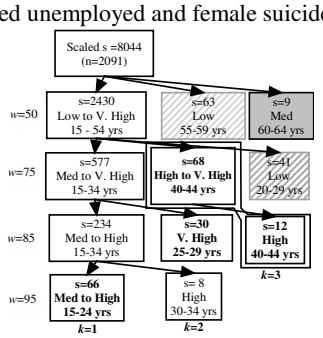
A. Bankruptcies number and female suicide rate



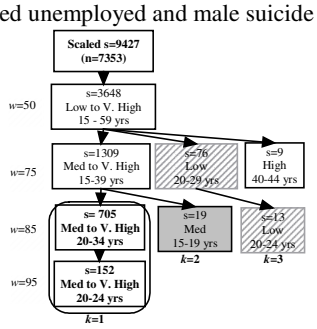
B. Bankruptcies number and male suicide



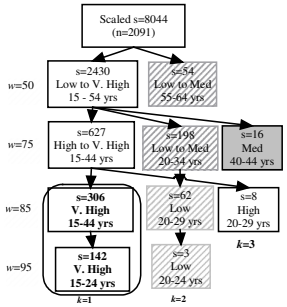
C. Registered unemployed and female suicide rate



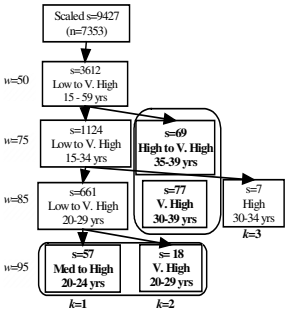
D. Registered unemployed and male suicide rate



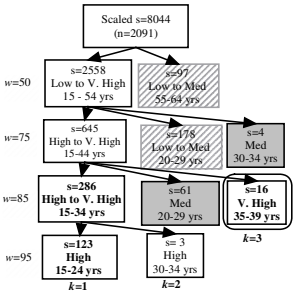
E. Divorce number and female suicide rate



F. Divorce number and male suicide rate



G. Orphan and unsupervised beneficiaries and female suicide rate



H. Orphan and unsupervised beneficiaries and male suicide rate

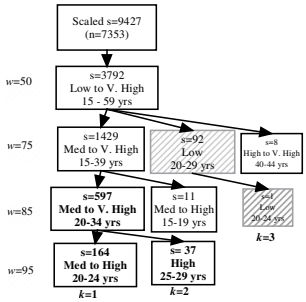


Figure 4. The *K*-MSA investigation on the suicide rate and social factors. Note that the shaded cells indicate the *low* or *med* social factor levels. Major findings or similar findings are circled together.

4.2. Bankruptcies

For both sexes, the 35-44 age group associates with the *very high* bankruptcy rate, but this association is stronger for females than males, as female maximum subarrays are found at lower K with large s values (k=2, w=75, s=110 and k=2, w=85, s=50) than male (k=3, w=75%, s=35 and k=3, w=85%, s=10), shown in Figure 4-A and B. Both sexes, in age groups below 35 years old, associate with *medium* to *high* bankruptcies; for females, the age group is 15 to 34 years old, and for males, 20-34 years old (both at k=1, w=85).

Furthermore, the maximum subarray at k=1 and w=95 shows a reasonably high s value (s=81) for female, suggesting a reasonably strong association of a young teenage female age group, 15 to 19, and *very high* bankruptcies. For males, the 20-24 age group, five years older than for females associated with *medium* to *high* bankruptcies and shows the highest s values (s=200) among all factors at the same threshold (k=1 at w=95), but when the age range becomes 25-29 years old, it associates with the high bankruptcies (k=2 and w=95), but its association is much smaller (s=28). Overall, it can be said that the very high bankruptcy rate may impact on females, especially teenager, compared with the early 20s males (as they associated with *med* to *high* bankruptcies levels).

4.3. Registered unemployed

Association of unemployment rate differs among sexes and their age, shown Figure 4-C and D. Firstly, a narrow older female age group, 40-44, associates with more than *high* registered unemployed rate (at k=2, w=75, s=68 and k=3, w=85%, s=12). Secondly, a large female age group, 15-34, is detected with *med* to *high* unemployed rate (at k=1, w=85), and this group separates out into two groups at further w values; 25-29 years old with *very high* (k=2, w=85) and 15-24 years old with *med* to *high* (k=1, w=95), shown in Figure 4-C. However, males respond differently; the older age group, 40-44, associated with *high*, but this is less significant (smaller s; s=9, k=3, w=75) than for females. On the other hand, a younger male age group, 20-34, associated with more than *med* unemployment rate significantly (large s; s=705, k=1, w=75) that further separates into the age group, 20-24 with more than *med* unemployed rate (k=1, w=95), shown in Figure 4-D.

According to the New Zealand census linked data, those experiencing unemployment had odds of subsequent suicide that were over 2.5 times higher than for those employed (Beautrais, 2003b). From this, in particular, early 40s and late 20s females associate with the extreme unemployed rate, but the same age groups for males are not significantly associated, whereas the mid male age group (20-34) associates significantly with the change of unemployment rate of larger than medium.

4.4. Number of divorces

Generally, the most extreme levels of divorce social factor are detected to associate with the very high suicide rate for both sexes, shown in Figure 4-E and F. However, the age groups between sexes are different with the divorce number. Females show the wider age range, 15-44, associating significantly with *very high* (large s; s=306 at k=1, w=85). Furthermore, a younger female age group, 15-24, with *very high* divorce rate is also detected with high s value (s=142 at k=1, w=95). For males, interestingly, an independent maximum subarray that does not belong to any other subarray is detected at k=2 and w=85 for 30s with *very high* divorce rate. Also, the maximum subarray at k=2 and w=75 shows the association of the late 30s. Furthermore w values, another two younger male age groups are detected; the male age group, the early 20s male with *med* to *high* (at k=1, w=95) and all 20s with *very high* divorce rate (at k=2, w=95).

In adult, it is more likely for people who have committed suicide than the general population to have never married, to be divorced or widowed, and to live alone (Beautrais, 2003b). While the male suicide rate is generally higher than the female rate (MOH, 2006), the sum of each maximum subarray also is higher for male than female. However, only divorce factor for female shows about three times higher sum of the maximum subarray (s=142) than males (s=57) at the same K and w position (k=1 and w=95 in Figure 4-G), when the female age range is 15-24 years old and is associated with *very high*.

From here, it could be possible that females respond to very high divorce rate, but not only adults: even younger females, from 15 to 24 as well as up to 44 years old. However, for males, generally high divorce rate associates with suicide in the 20s and 30s.

4.5. Orphans and unsupervised child beneficiaries

From the larger maximum subarrays (at k=1, w=85), both sexes show similar trends (Figure 4-G and H). However, the female age group starts five years younger than for males; females, 15-34 and males, 20-34, associate with more than *high* and

med to *high* respectively. Furthermore, this large age group is divided into specific levels. For females, the age group, 15-24, associates with *high* (k=1, w=95 with a reasonably high suicide rate (s=123). Besides, another older female age group, 35-39, is detected with *very high* orphan numbers (k=3, w=85 and s=16), though is not significant as a small s value. Similar to female, two male groups are separated out at w=95 from the larger maximum subarray at w=85. Firstly, the age group, 20-24, with *med* to *high* and has a reasonably high suicide rate (large s value; s=164), and secondly, the age group, 25-29, with *high* orphan numbers (but smaller s value; s=37).

Overall, the high suicide rate (larger s values) at the high orphan number is detected for females in two age groups, 15-24, and 35-39 years old, but for males, generally, in their 20s or perhaps up to 34 years old.

5. CONCLUSIONS

The K-MSA successfully detects the range of the threshold criteria to describe the maximum associations of the suicide rate, suicide age groups and selected social factors. Additionally, changing the weight parameter setting successfully provides detailed associations for the specific suicide range and social factor levels.

Generally, the similar association pattern for bankruptcies and unemployment, and divorce and orphan numbers are found from the suicide rate for females and males, respectively. However, these associations differ between the sexes. The maximum suicide age groups are found to be wider for females than males (females, 15 to 64 years old; and males, 15-59 years old). Among four selected social factors, especially, an extremely high divorce number is the most significant factor to associate with the high suicide rate for both sexes. However, the female suicide age group starts much younger, from 15 to 44 years old, compared with males, where it is generally in the 20s to 30s. Knowing that the New Zealand male suicide rate is generally higher than the female rate, the sum of each maximum subarray is generally found to be larger for males than females at the same K and w values. However, a single higher maximum sum is found for females between 15 and 24 years old against a *very high* divorce rate, suggesting that *very high* divorce rate is not only affecting adults, but also younger people, particularly for females. While the maximum subarray in this study is designed to keep the time sequence of the suicide occurrence and the social factor level, the extremely high social factor level may impact on the suicide rate time-sensitively as they happened in the same year.

An experimental investigation using the K-MSA and a new approach, changing the weight parameter, successfully provide detailed associations for the specific suicide age groups and social factor levels. Results do not identify the direct causal factors for the individual suicidal behaviour, but represent the national trends. In future, different social factors can capture background information for suicidal behaviour, and the time sequence can be a moving average or time lags to allow highlighting a long-term association between suicide and changing social factor levels. The K-MSA can be an alternative data mining tool to provide knowledge about the data set to help minimising the future suicide rate by setting up a specific policy for the specific age groups for different sexes.

6. ACKNOWLEDGMENTS

Our thanks to P. Pearson for processing data and helping with editorial work.

7. REFERENCES

Bae, S.E. and T. Takaoka (2005), Improved algorithms for the K-maximum sub-array problem for small K. In *Proc. of 11th Intl. Computing and Combinatorics Conference (COCOON '05)*, Kunming, China, 621-631.

Beautrais, A. (2003a), Suicide in New Zealand I: time trends and epidemiology, *The New Zealand Medical Journal,* 116, No. 1175.

Beautrais, A. (2003b), Suicide in New Zealand II: a review of risk factors and prevention, *The New Zealand Medical Journal,* 116, No. 1175.

Fukuda, K. and T. Takaoka (2007), Analysis of Air Pollution (PM_{10}) and Respiratory Morbidity Rate using K-Maximum Sub-array (2-D) Algorithm, *In Proc. of the 2007 ACM Symposium on Applied Computing* (SAC'07), Seoul, Korea, 153-157.

Ministry of Health (2006), New Zealand Suicide Trends, Mortality 1921-2003 Hospitalisations for intentional self-harm 1978-2004, *Public Health Intelligence monitoring Report No. 10,* Ministry of Health, Wellington.