

On Estimate of Malaysian Mortality Rates Using Interpolation Methods

Nur Idayu Ah Khaliludin*, Zarina Mohd Khalid and Haliza Abd.Rahman

Department of Mathematical Sciences, Universiti Teknologi Malaysia
81310 UTM Johor Bahru, Malaysia

*Corresponding author: ayu.khalil@gmail.com

Article history

Received: 18 March 2019

Received in revised form: 18 March 2019

Accepted: 20 May 2019

Published online: 1 August 2019

Abstract Life table is a table that shows mortality experience of a nation. However, in Malaysia, the information in this table is provided in the five-years age groups (abridged) instead of every one-year age. Hence, this study aims to estimate the one-year age mortality rates from the abridged mortality rates using several interpolation methods. We applied Kostaki method and the Akima spline method to five sets of Malaysian group mortality rates ranging from period of 2012 to 2016. The result were then compared with the one-year mortality rates. We found that the method by Akima is the best method for Malaysian mortality experience as it gives the least minimum of sum of square errors. The method does not only provide a good fit but also, shows a smooth mortality curve.

Keywords Actuarial science; life table; spline interpolation; mortality; expansion of life table.

Mathematics Subject Classification 65B10, 62-07

1 Introduction

Life table which is also known as mortality table provides the survivorship and mortality experience of a population be it a group of cancer survivor or an animal population. The information included in the table (but not necessarily all) are the instantaneous mortality rates, the probability of dying and living as well as the expected years to live. This table is used mainly by actuaries in the calculation of annual provision of life insurance and Family Takaful as well as the future value of the retirement fund. They also need the life table to compute the monthly premiums and contributions charged to the policyholders. Furthermore, in the presence of the new accounting regulation Malaysian Private Entities Reporting Standard (MPERS), all the government agencies such as the public universities are required to evaluate the net present value of employee benefits using the actuarial method which made the needs of the mortality table become more imperative.

There are two types of life table; abridged and complete life table. The complete life table shows the population information for every single age (0,1,2,3,4,5,...) whereas the former, the

abridged life table shows the mortality experience for the age group which is usually a group of five-years ages (0-4,5-9,10-15,...) until a certain maximum age. In Malaysia, Unfortunately, producing a complete life table is tedious work as it requires research on death rates and advanced computational methods. The number of deaths is concentrated heavily during the first year of life and at oldest age particularly after age 65 whereas, at the other ages, the number of deaths had increased gradually. Hence, the mortality data is usually pooled in the age groups to reduce the variability of deaths and as a consequence, there are only abridged life table available publicly in Malaysia.

Researchers, however, are able to estimate the single age mortality rate from the abridged life tables using simple interpolation methods. In this paper, we aim to determine the best method to estimate the single age mortality rate from the grouped or abridged life table.

1.1 Review of Existing Method

One of the earliest approaches to obtain the one-year age mortality rates from the abridged rates is the relational methods. Relational methods utilize the available complete mortality table to estimate the mortality rates for all ages from the abridged table and thus, allow the mortality comparison between a population with that of other population, or between past years mortality experience with the current experience. The relational methods are often favoured by the government institutions as they are the shortest and simplest methods of constructing the mortality table. The most renowned relational method is the Brass logit method which relates two mortality pattern using only two known data points [3]. The Brass method was applied extensively in demography areas such as to study the population in New Zealand and Malaysia [4, 8]. Nevertheless, it should be mentioned that Brass did not aim at perfect goodness-of-fit, but used a simple method to obtain decent approximations. A modification of this method is done by Kostaki where he improved the Brass logit method by incorporating more mortality data in the model [6]. The Kostaki method reduces the dependency to the reference life table since the data for each five-year interval of the abridged life table are paired with those from the reference life table in the same interval as compared to the Brass method which depends on only two data from the abridged life table.

Alternatively, the abridged life table can also be interpolated to obtain the complete life table. Initially, the interpolation methods discovered were not intentionally used for the expansion of the abridged table, but are more into the numerical methods area for instances, Lagrange polynomials. This method was applied by Elandt-Johnson and Johnson to estimate the numbers of survivors, l_x that are not tabulated in the abridged life tables as a linear combination of six particular functions of age. He called this method the six-term Lagrangian interpolation method [5].

Another interpolation method is the spline interpolation. This method is considered to be the most recent yet complicated method. Spline provides a smooth curve to the data by adjusting the degree of the polynomials which is an outstanding feature compared to the other methods mentioned before. The only spline that is used in this process is Akima spline [1].

In this paper, we present the comparative analysis of two methods namely the Akima and Kostaki method in estimating the one-year age mortality rates from the group mortality rates and evaluate the performance of each method using the mean of squares error (MSE). This paper is organized into four sections. Section 2 elaborates the methodology adopted to

estimate the one-year age mortality rates whereas Section 3 shows the results and discuss the findings. Finally, we conclude our research in Section 4.

2 Methodology

In this section, the selected methodology to estimate the complete mortality rates from the abridged mortality rates are presented.

2.1 Data

This research requires the data on the grouped mortality rates from the abridged life tables and the one-year age mortality rates. The grouped mortality rates are fundamentally the accumulated one-year age mortality rates. They were supplemented by the Department of Statistics of Malaysia for a five year period from the year 2012 to 2016, disaggregated by gender and 18 age groups.

2.2 Kostaki Method

Kostaki gives a simple non-parametric method pairing each five-year interval of the abridged mortality rates with the the one-year age mortality rates in the same interval.

The idea of this method is that in every age interval $[x, x + n)$, the group force of mortality μ_x is a constant multiple of the one-year age force of mortality ${}^s\mu_x$ in the same age interval. We thus estimate a constant ${}_nM_x$ for each age interval $[x, x + n)$ using equation (2).

$${}_nM_x = \frac{\ln(1 - {}_nm_x)}{\sum \ln(1 - {}^sm_x)} \quad (1)$$

Where ${}_nm_x$ is the group mortality rate for the interval $[x, x + n)$ and sm_x is the one-year age mortality rate at age x , with $x \in [x, x + n)$.

The complete mortality rates m_x are calculated as follows:

$$m_x = 1 - (1 - {}^sm_x)^{{}_nM_x} \quad (2)$$

using

$$\begin{array}{lll} {}_4M_1 & \text{for} & x \in [1, 4) \\ {}_5M_5 & \text{for} & x \in [5, 9) \\ & & \cdot \\ & & \cdot \\ & & \cdot \\ {}_5M_{75} & \text{for} & x \in [75, 80) \end{array}$$

2.3 Akima Method

Akima method is based on a piecewise function composed of a set of polynomials, each of degree three, at most, and applicable to successive interval of the given points. In this method, the slope of the curve is determined at each given point locally, and each polynomial representing a portion of the curve between a pair of given points is determined by the coordinates and the slopes at the points.

In order to estimate the one-year mortality rates at age i , say m_i , the method uses four sets of four consecutive data points that include m_i i.e, $(m_{i-3}, m_{i-2}, m_{i-1}, m_i)$, $(m_{i-2}, m_{i-1}, m_i, m_{i+1})$, $(m_{i-1}, m_i, m_{i+1}, m_{i+2})$, $(m_i, m_{i+1}, m_{i+2}, m_{i+3})$. The formula is:

$$m_i = m'_{i1}w_{i1} + m'_{i2}w_{i2} + m'_{i3}w_{i3} + m'_{i4}w_{i4} \quad (3)$$

where m'_{is} is the first derivative of the s th consecutive data points ($s = 1, 2, 3, 4$) and w_{is} is the weight respective to the s th consecutive data points. m'_{is} and w_{is} is further explained by the equation (4) and (5) respectively.

$$\begin{aligned} F(i, j, k, l) = & [(m_j - m_i)(x_k - x_i)^2(x_l - x_i)^2(x_l - x_k) \\ & + [(m_k - m_i)(x_l - x_i)^2(x_j - x_i)^2(x_j - x_l)] \\ & + [(m_l - m_i)(x_j - x_i)^2(x_k - x_i)^2(x_k - x_j)] \\ & / [(x_j - x_i)(x_k - x_i)(x_l - x_i)(x_k - x_j)(x_l - x_k)(x_l - x_j)] \end{aligned} \quad (4)$$

In the equation (4), m_i, m_j, m_k and m_l , are the mortality rates of four consecutive data points that include the rate we want to estimate and x_i, x_j, x_k and x_l are the sets of four consecutive ages of the mortality rate we want to estimate.

Also, equation (5) which defined the weight w_{is} is the one over product of the sum of square of deviations from a straight line of the least-square fit and the distance factor.

$$w_{i,j,k,l} = \frac{1}{\sum(y - \beta_0 + \beta_1 x)^2 \times [(x_j - x_i)^2 + (x_k - x_i)^2 + (x_l - x_i)^2]} \quad (5)$$

$\sum(y - \beta_0 + \beta_1 x)^2$ is the sum of square of deviations where y is the observed mortality rate and β_0 and β_1 are the coefficients of the least square fit to our mortality data.

2.4 Evaluation of the Methods

The result calculated must be evaluated to decide which of the two methods are the most suitable for the Malaysian mortality data. We assessed the error of each method using the mean of sum of square of the deviations (MSE) between the fitted value and the empirical observations. The method that produced lesser errors is considered as the best method.

$$\text{MSE} = \frac{\sum(\hat{m}_x - m_x)^2}{n}, \quad (6)$$

where n is the total number of age, \hat{m}_x is the estimated rate obtained from the fitted model and m_x is the observed one-year mortality rate.

3 Results and Discussion

In this research, five sets of abridged and one-year mortality rates were obtained spanning from the period of 2012 to 2016 and separated by gender.

Certain features of mortality rates are immediately apparent in Figure 1 and Figure 2. As expected, the variability across ages is high especially in the first thirty years of life which shows wavering mortality curves. Mortality rates drop sharply in the first few years of life especially prominent between ages 0 and 1. This is because infants have weak immune systems and are very susceptible to all kind of diseases. From around age 10 onwards, mortality rates increase, with a step increase in late teenage years, especially outstanding for males and attributable to a higher death rate from accidents. This phenomenon is called the accident hump. It also evident from Figure 1 and Figure 2 that the variability between empirical mortality rates across the five years of observation is much lesser compared with the variability between gender and across ages. We initially expect the mortality rates to drift downwards over time but this drifts are weakly visible across the years 2012 to 2015 much less for males than females. Even so, the female rates are unstable around ages 15 to 40 where the rates in 2016 shifted more to the left.

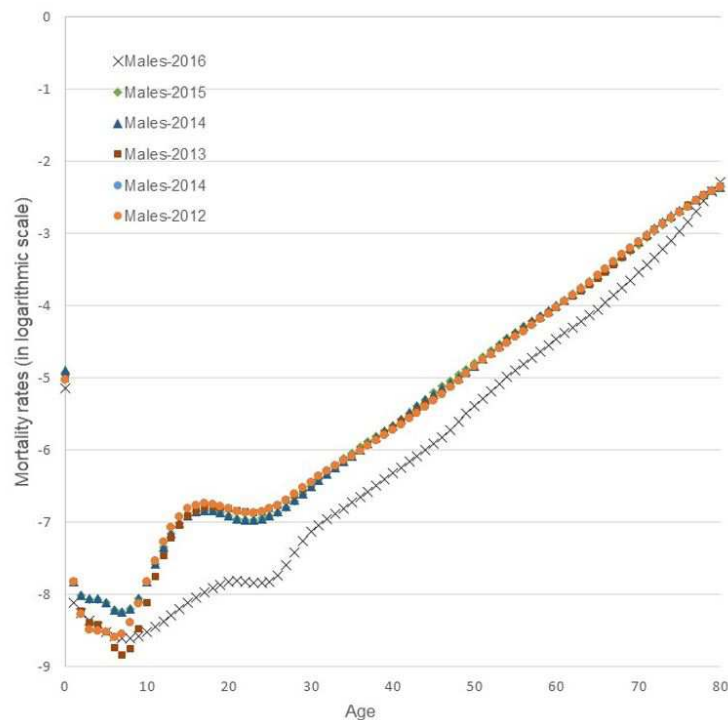


Figure 1: Empirical Mortality Rates for Males (2013-2016)

Figure 3 combined all the graphs for all years and genders. It shows that males have higher mortality rates compared to the females except in the year 2016. 2016 recorded the lowest male mortality rates while the rates follow closely to the female rates which indicates that the male mortality will eventually improves in the future and the gender difference will be less significant. Both male and female rates finally converge in the later years around ages 80.

We then applied the Kostaki and Akima method to the set of abridged mortality rates and obtained the result as shown is the Figure 4 and Figure 5.

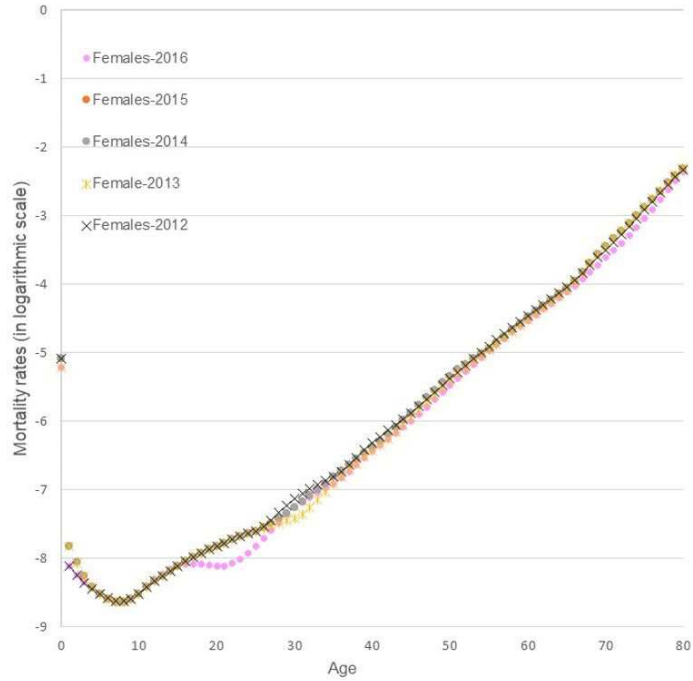


Figure 2: Empirical Mortality Rates for Females (2013-2016)

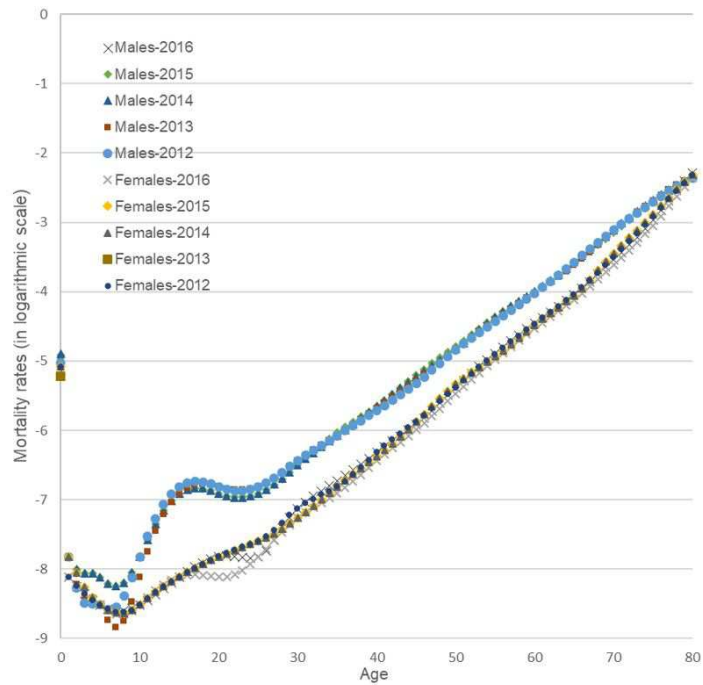
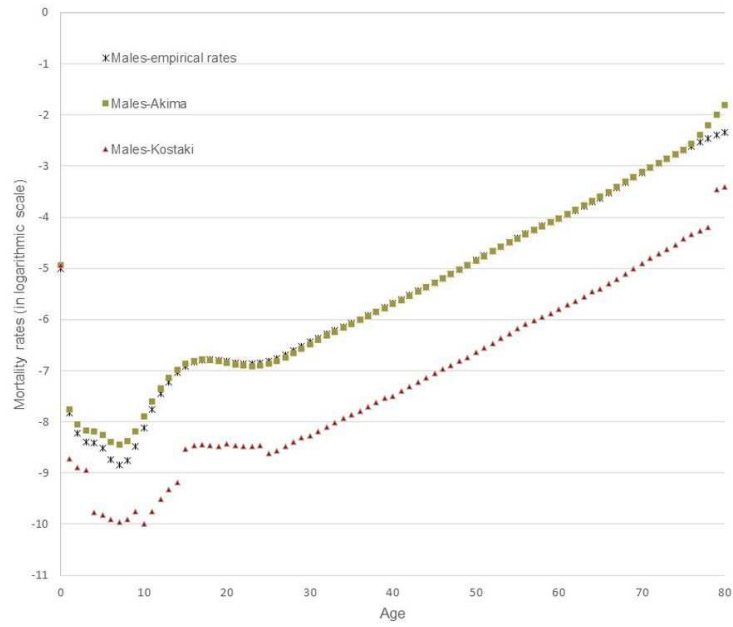
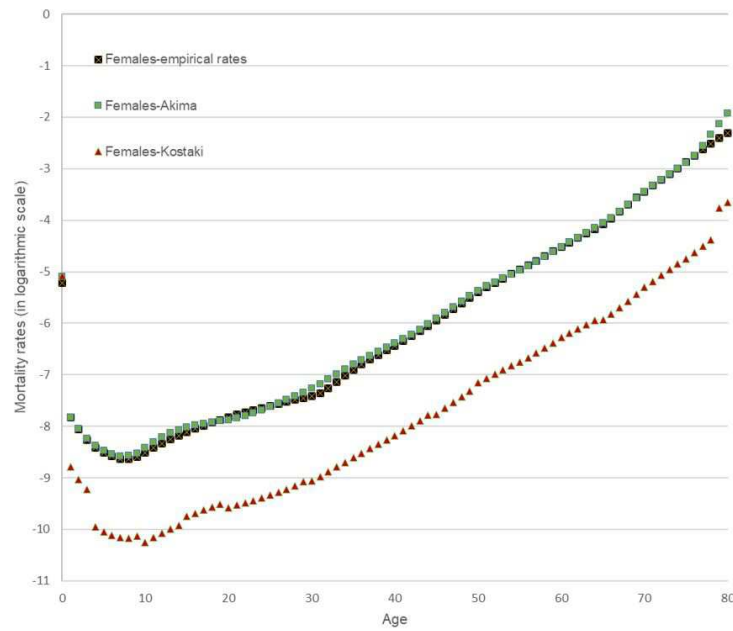


Figure 3: Empirical Mortality Rates (2013-2016)



(a) Male



(a) Female

Figure 4: Empirical Mortality Rates for Year 2013 Together with Rates Estimated by Akima and Kostaki Methods

Figure 4 presents the empirical mortality rates for the year 2013 together with the rates estimated by Akima and Kostaki methods. Akima produces a smooth mortality curve which is very similar to the empirical rates. On the other hand, Kostaki method fails to estimate the mortality rates for both genders and the mortality curve produced by this method is not smooth although a similar one-year mortality rates was used. This is because the Akima method solely depends on the empirical abridged rates whereas the Kostaki method relies heavier on the other rates. Since, our data ends at age group 80 and both methods require the mortality rates of age before and after the desired rate, a huge variance appears around age 80 where all the produced rates are higher than the empirical rates.

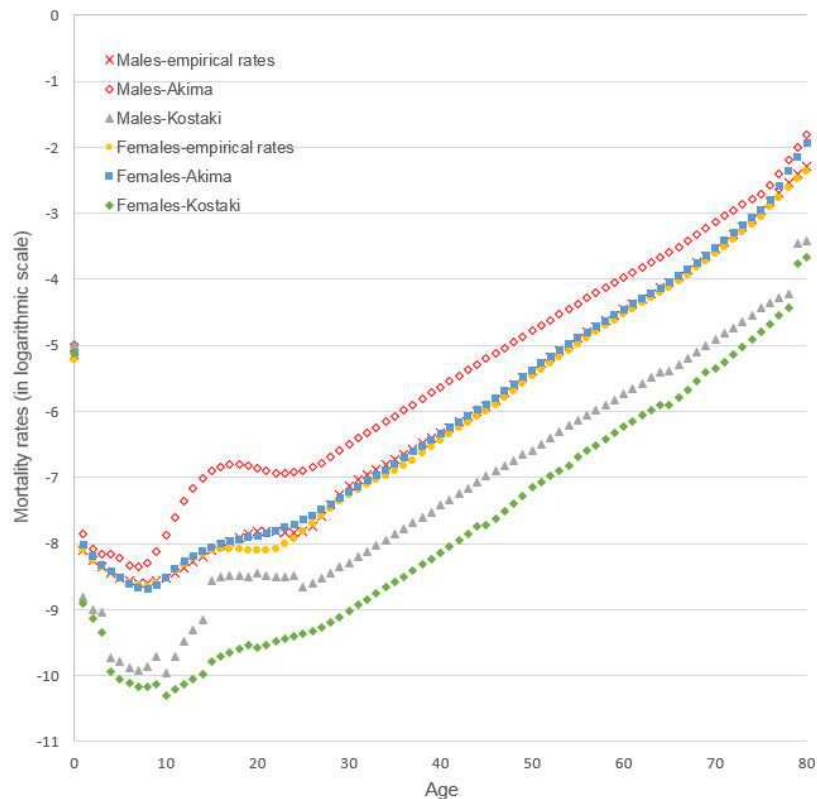


Figure 5: Empirical Mortality Rates for Year 2016 Together with Rates Estimated by Akima and Kostaki Methods

On the other hand, Figure 5 above shows that both Kostaki and Akima methods fail to obtain a good approximation of one-year age mortality rates for males in 2016. The female rates estimated using the Akima method shows a good fit though and the rates are actually the same as the male mortality rates. This is not surprising as we have observed that there is not much gender differential in the 2016 mortality rates.

We then assessed the error of both methods using the mean of square of error to attain a clearer result. Table 1 and Table 2 display the error for each covered years for males and females respectively. As what we have anticipated, the Akima method provides lesser error than the Kostaki methods. Nevertheless, the errors are more significant over the years.

Based on the result obtained, we can conclude that Akima is the best method to estimate the one-year age mortality rate.

Table 1: Sum of Squares of the Absolute Deviations Between the Fitted Value and the Empirical Observations (Males)

Methods	2012	2013	2014	2015	2016
Kostaki	0.00049	0.00049	0.00056	0.00050	0.00031
Akima	0.00008	0.00008	0.00008	0.00008	0.00011

Table 2: Sum of Squares of the Absolute Deviations Between the Fitted Value and the Empirical Observations (Females)

Methods	2012	2013	2014	2015	2016
Kostaki	0.00039	0.00040	0.00039	0.00037	0.00030
Akima	0.00004	0.00003	0.00003	0.00004	0.00005

4 Conclusion

This research aims to estimate one-year mortality rates from five-years age mortality rate using Akima and Kostaki methods. We identify that the Akima method produces a better set of one-year mortality rates for Malaysia although the Kostaki method is capable to estimate the one-year mortality rate, it still cannot produce a smooth mortality curve. Nevertheless, researchers must be cautious in selecting between the smoothness of curve and the goodness of fit as producing a smooth curve may eliminate the underlying value of the mortality rates. However, the downside of the Akima method is that it cannot estimate the oldest age mortality rate particularly after ages 75. Thus, it is suggested that future researchers improve the method so that it is more accurate for the old ages. Researchers can also perhaps complete this method with the extrapolation methods.

Acknowledgments

The author would like to thank the Ministry of Education Malaysia and GE STEM grant with vot no. 07397.

References

- [1] Akima, H. A method of univariate interpolation that has the accuracy of a third degree polynomial. *ACM Transactions on Mathematical Software*. 1991. 17 (3): 341-366.
- [2] Baili, P., Micheli, A., Montanari, A., and Capocaccia, R. Comparison of four methods for estimating complete life tables from abridged life tables using mortality data supplied to EURO CARE-3. *Mathematical Population Studies*. 2005. 12(4): 183-198.
- [3] Brass, W. (1971). Mortality models and their uses in demography. *Transactions of the Faculty of Actuaries*. 1971. 33: 123-142.
- [4] Department of Statistics of Malaysia. *Abridged Life Tables, Malaysia, 2015-2017*. Malaysia: Department of Statistics, Malaysia. 2017

- [5] Elandt-Johnson, R. and Johnson, N. *Survival Models and Data Analysis*. New York: John Wiley. 1980.
- [6] Kostaki, A. A relational technique for estimating the age-specific mortality pattern from grouped data. *An International Journal of Mathematical Demography*. 2000. 1(9): 83-95.
- [7] Kostaki, A. and Panousis, V. Expanding an abridged life table. *Demographic Research*. 2001. 5(1): 1-22.
- [8] Statistics New Zealand. *New Zealand Period Life Tables: 2012-14*. Wellington: Statistics New Zealand. 2015.