

# Mortality Rates Estimation Using Whittaker-Henderson Graduation Technique

MARIELYNN E. CHANCO  
Social Security System  
Quezon City, Philippines  
*marielynn\_chanco21@yahoo.com*

## Abstract

This study aims to find a good estimate of mortality rates of the Philippine Social Security System members exposed for ten years. The assumptions and process of establishing the exposure and number of deaths in the population were briefly discussed in this paper. The estimation of initial mortality rates were done by getting the ratio of deaths and exposure obtained based on the assumptions. Whittaker-Henderson graduation technique was applied to easily maintain the smoothness as well as the fitness of the graduated mortality rates relative to the crude mortality rates. Different values are tested for the parameters in the Whittaker-Henderson graduation formula which corresponds to the smoothness and fitness to estimate the graduated mortality rates. Graduated mortality rates were further adjusted using partial credibility theory to obtain the ideal behavior of the mortality for older ages.

**AMS Classification:** 62P05, 91G70, 97M30

**Key words:** Actuarial Mathematics, mortality rates, survival models

## 1 Introduction

Mortality rate is the probability of dying within a specified period of time. It is usually calculated by getting the ratio of total number of deaths and exposure. Exposure pertains to the annual number of units of human life which are subject to death, disability, or some other decrement, within a defined period of observation.

Mortality rates estimation is crucial for population projections as well as other related projections such as benefits and contributions for insurance companies. Several mortality studies and models are already established for specific demographic and economic scales in developed countries. In the Philippines, private insurance companies usually depend on published mortality rates made by international statistical organizations to estimate their own mortality rates.

Data from the Republic of the Philippines Social Security System was used for the mortality rates estimation for years 2003 up to year 2012. Annual observations of deaths and active lives were used in the determination of exposure and raw mortality rates. Exposure was estimated using uniform distribution assumption within one year for ten years of observation period. Calculation of exposure and raw mortality rates were further discussed in the methods and procedure section.

Raw mortality rates (crude mortality rates) were then smoothed and fitted using Whittaker-Henderson graduation technique. It is one of the most commonly used smoothing methods

for crude mortality rates. It has the capacity to explicitly balance the smoothness as well as the fitness of the smoothed curve relative to the crude rates through its parameters by minimizing a difference equation.

Partial credibility theory was used to further adjust the graduated mortality rates using the Whittaker-Henderson graduation technique to obtain the desired curve behavior for alder ages. Credibility theory is commonly used to deal with the randomness of the data caused by the adjustments made in the crude mortality rates using the actual experience or other historical information that are much more credible than the actual graduated rates.

## 2 Methods

Exposure refers to the annual number of units of human life which are subject to death, disability, or some other decrement, within a defined period of observation. This will be used as the denominator in finding the mortality rate or rate of decrement due to some other cause. The numerator is the number of observed deaths or the number of lives that left the defined group due to the cause being measured.

For the distribution of deaths over a unit age interval, age  $x$  at death is assumed to be the difference between the calendar year of birth (CYB) and calendar year of death (CYD),

$$x = CYD - CYB$$

The occurrence of deaths is assumed to be linear over the age interval  $[x, x + 1]$ , thus it is uniformly distributed over a unit age interval.

Members are classified according to how they enter or exit the group under observation within a specific time interval

- Starters – members who are active upon the commencement of the observation period
- New entrants – members who entered the group during the observation period
- Withdrawals – members who ceased to participate during the observation
- Deaths – members who died during the observation period
- Enders – members who are active through the end of the observation period

Exposure is denoted by  $E_x$ .

Let  $l_x^y$  be the number of lives of age  $x$  living at the end of calendar year  $y$ ,  $m_x^y$  the number of migrants (new entrants less withdrawals) of age  $x$  at the end of calendar year  $y$ , and  $d_x^y$  the number of deaths of age  $x$  at the end of calendar year  $y$ . Then

$$l_{x+1}^{y+1} = l_x^y + m_x^{y+1} - d_x^{y+1} \quad (1)$$

Since deaths are assumed to be uniformly distributed over a unit age interval, it is assumed that migrants are also uniformly distributed. Hence, the exposure,  $E_x$ , over a unit age interval is given by

$$E_x = l_x^y + \frac{1}{2}m_x^{y+1} \quad (2)$$

Using (1), we can rewrite (2) as

$$E_x = l_x^y + \frac{1}{2} \left( l_{x+1}^{y+1} - l_x^y + d_x^{y+1} \right) = \frac{1}{2} \left( l_x^y + l_{x+1}^{y+1} + d_x^{y+1} \right) \quad (3)$$

Assuming the beginning of the calendar year as  $y + 1$  (end of calendar year  $y$ ) and the end of calendar year as  $y + 3$ , the total exposure is

$$E_x = \frac{1}{2} (l_x^{y+1} + l_x^{y+2}) + \frac{1}{2} (l_{x+1}^{y+1} + l_{x+1}^{y+2} + l_{x+1}^{y+3}) + \frac{1}{2} (d_x^{y+1} + d_x^{y+2} + d_x^{y+3}) \quad (4)$$

and in general,

$$E_x = \frac{1}{2} \sum_{i=y}^{y+m-1} l_x^i + \frac{1}{2} \sum_{i=y}^{y+m} l_{x+1}^i + \frac{1}{2} \sum_{i=y}^{y+m} d_x^i \quad (5)$$

where  $m$  is the year count up to the end of the valuation period.

The mortality rate  $q_x$  is a measure of the number of deaths over the exposed lives with age  $x$  over a given time interval. It is given by  $q_x = d_x/E_x$ , where  $d_x = \sum_{i=y}^{y+m-1} d_x^i$ . After the estimation of mortality rates, mortality rates are graduated using the Whittaker-Henderson graduation method. In the formula used,  $M$  is minimized and the variable  $\hat{q}_x$  is optimized using an Excel<sup>®</sup> solver. The formula goes as follows:

$$M = \sum_{x=1}^n w_x (q_x - \hat{q}_x)^2 + h \sum_{x=1}^{n-z} (\Delta^z \hat{q}_x)^2 \quad (6)$$

where  $z = 2, 3, 4, 5$  (whichever gives the best fit),  $h = 10, 50, 100$  (for both males and females),  $x = 15$  to 100 years of age,  $q_x$  is the crude mortality rate,  $\hat{q}_x$  is the graduated mortality rate with initial value set to  $q_x$ ,  $\Delta^z \hat{q}_x = k \cdot \hat{q}_x$ ,

$$w_x = \frac{e}{\sum_{x=1}^n E_x} E_x \quad (7)$$

and  $k$  is an  $(n - z) \times n$  matrix containing coefficients of the polynomial corresponding to a chosen degree,

$$k = \begin{pmatrix} B & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & B \end{pmatrix}$$

with

$$B = \begin{cases} [1, -2, 1] & \text{if } z = 2 \\ [-1, 3, -3, 1] & \text{if } z = 3 \\ [1, -4, 6, -4, 1] & \text{if } z = 4 \\ [-1, 5, -10, 10, -5, 1] & \text{if } z = 5 \end{cases}$$

The first term represents an expression for the goodness of fit of the graduated values to the raw rates; the second term is an expression for the degree of smoothness of the graduated values. The factor  $h$  is determined arbitrarily. A low value of  $h$  puts more emphasis on fit than smoothness and returns the raw rates; there is essentially no graduation. A high value of  $h$  puts more emphasis on smoothness than fit and yields a least squares fit to a polynomial of degree  $n - 1$ . The method calculates directly a complete set of graduated values which achieve the desired balance between fit and smoothness. Also the total number of deaths and the average age at death, when the exposures are used as weights, will be the same for the graduated rates as for the raw rates.

The actual and graduated data on mortality rates have an increasing trend from age 15 to age 80 for males and up to age 86 for females, then gradually decrease at the older ages.

This behavior tends to contradict the natural phenomenon that mortality rates for older ages shall be higher than the younger ages. Thus, adjustments are made to come up with a realistic trend at the older ages. The basis for this trend was the trend of the published mortality rates from previous SSS Actuarial Valuations. In determining the trend of the previous mortality rates, the *logit* function was applied. A *logit* is a unit of measurement which reports relative differences between estimates. Logits are equal interval levels of measurement, which means that the distance between each point on the scale is equal. Applying the *logit* function to the previous mortality rates would then result in a series of linear trends. From these resulting trends, the expected behavior of new mortality rates can be assumed to follow the same trend, thus expected mortality rates could be determined. The expected mortality rates are needed in adjusting the graduated mortality rates using credibility theory which will be discussed later in this paper. Using the published mortality rates in the previous valuation years 2003, 2007 and 2011, the *logit* for age  $x$  in each of the valuation year  $t$  is calculated as follows:

$$\text{logit}_x^t = \log \frac{\hat{q}_x}{1 - \hat{q}_x} \quad (8)$$

For the resulting curve for each year, a linear function  $y = mx + \bar{b}$ ,  $m, \bar{b} \neq 0$ , is fitted for ages 15–54. Coefficients  $m$  and  $\bar{b}$  were initially set to arbitrary nonzero values to compute for initial  $y$  values. The square difference of  $y$  and  $\text{logit}_x^t$ , or  $L = (y - \text{logit}_x^t)^2$ , was then determined and minimized using an Excel<sup>®</sup> solver to obtain the optimal values of  $m$  and  $\bar{b}$  that would give the best fit.

Due to the nonlinear curve that resulted from the application of logit function at ages 55–99, a cubic polynomial function,  $y = ax^3 + bx^2 + cx + d$ , for  $a, b, c, d \neq 0$ , was fitted instead. The coefficients were determined using the same method as with the fitting of the linear function. The optimal coefficients obtained from this procedure were denoted by  $m_{2003}$ ,  $m_{2007}$ ,  $m_{2011}$ ,  $\bar{b}_{2003}$ ,  $\bar{b}_{2007}$ ,  $\bar{b}_{2011}$ , for ages 15–54. Then for ages 55–99, the optimal coefficients are denoted by  $a_{2003}$ ,  $a_{2007}$ ,  $a_{2011}$ ,  $b_{2003}$ ,  $b_{2007}$ ,  $b_{2011}$ ,  $c_{2003}$ ,  $c_{2007}$ ,  $c_{2011}$ ,  $d_{2003}$ ,  $d_{2007}$ , and  $d_{2011}$ . The corresponding values for each of these optimal coefficients were then used to estimate the coefficients of the linear and cubic polynomial function which corresponds to the *logit* of the mortality rates for the 2014 valuation ( $m_{2013}$ ,  $\bar{b}_{2013}$ ,  $a_{2013}$ ,  $b_{2013}$ ,  $c_{2013}$  and  $d_{2013}$ ).

Alternatively, Excel<sup>®</sup> data analysis regression can be used to find the optimal coefficients in the polynomial functions mentioned earlier. The following equations were used to estimate the coefficients for ages 15–54:

$$\begin{aligned} m_{2013} &= \frac{1}{2} m_{2011} \left( \frac{m_{2007}}{m_{2003}} + \frac{m_{2011}}{m_{2007}} \right) \\ \bar{b}_{2013} &= \frac{1}{2} \bar{b}_{2011} \left( \frac{\bar{b}_{2007}}{\bar{b}_{2003}} + \frac{\bar{b}_{2011}}{\bar{b}_{2007}} \right) \\ y_{2013} &= m_{2013}x + \bar{b}_{2013}, \quad x = 15, 16, \dots, 53, 54. \end{aligned}$$

For ages 55–99, the following equations were used:

$$\begin{aligned}
 a_{2013} &= \frac{1}{2} a_{2011} \left( \frac{a_{2007}}{a_{2003}} + \frac{a_{2011}}{a_{2007}} \right) \\
 b_{2013} &= \frac{1}{2} b_{2011} \left( \frac{b_{2007}}{b_{2003}} + \frac{b_{2011}}{b_{2007}} \right) \\
 c_{2013} &= \frac{1}{2} c_{2011} \left( \frac{c_{2007}}{c_{2003}} + \frac{c_{2011}}{c_{2007}} \right) \\
 d_{2013} &= \frac{1}{2} d_{2011} \left( \frac{d_{2007}}{d_{2003}} + \frac{d_{2011}}{d_{2007}} \right) \\
 y_{2013} &= a_{2013}x^3 + b_{2013}x^2 + c_{2013}x + d_{2013}, \quad x = 55, \dots, 99.
 \end{aligned}$$

The adjusted  $q_x$  can be derived by reversing the *logit* function used earlier, given by

$$\text{adjusted } q_x = \frac{\exp(y_{2013})}{1 + \exp(y_{2013})} \quad (9)$$

To complete the construction of mortality rates, the graduated mortality rates were associated with the adjusted  $q_x$  using credibility weights. Credibility weights, whose values depend on the credibility of the observed data (i.e., credibility is 0 if data is too small, and 1 if data is large enough to be fully credible), are factors used in combining observed data with other information. With the graduated mortality rates as observed data and adjusted  $q_x$  as other information, the credibility weighted estimate is given by the following formula:

$$\text{estimated } q_x = \text{cred} \times \text{observed data} + (1 - \text{cred}) \times \text{other information} \quad (10)$$

where

$$\text{cred} = \sqrt{\frac{\text{average no. of claims from observed data}}{\text{expected no. of claims}}} \quad (11)$$

and

$$\text{expected no. of claims} = \left(\frac{z}{k}\right)^2 \left(\frac{\sigma}{\mu}\right)^2 \quad (12)$$

where  $z$  is the area under the standardized normal distribution corresponding to the credibility level (probability) chosen,  $k$  is the specified frequency value away from the mean,  $\sigma$  is the standard deviation of the observed data (using frequency rate), and  $\mu$  is the average frequency rate per exposure of the observed data.  $k$  is selected such that the number of claims observed falls within  $\pm k\%$  away from the expected number of claims.

### 3 Results

Mortality data were obtained from Social Security System (SSS), Philippines. As commonly used in applications, a constant value of 1,082 claims was used as the expected number of claims which corresponds to a 90% credibility and random error of at least 5%. It was assumed that there was a 90% chance of being within 5% of the mean, thus, there was also a 5% chance of being outside on either tail, for a total of 10% probability of being outside the acceptable range, if there was no variation in the size of claim. The number of expected claims required for full credibility was increased by a factor of  $\sigma^2/\mu^2$  if there was variation in the claim severity.

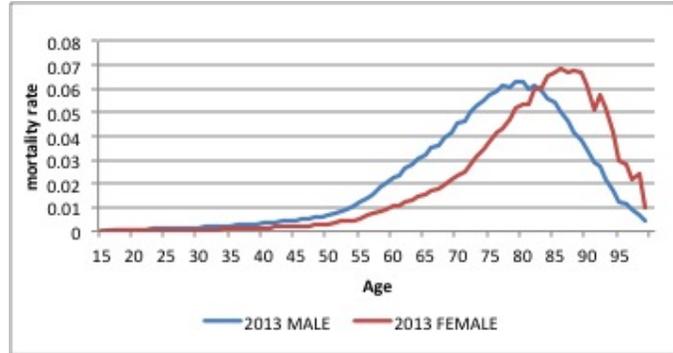


Figure 1: Crude mortality rates of males vs. females for the year 2013.

The credibility weighted estimates represented mortality rates for year 2013 denoted by  $q_x^{2013}$  were used as base rates in the projection and construction of the mortality table.

Figure 1 shows the crude mortality rates of male vs female for end of year 2013. The male population had higher rate of mortality than the female population for ages 15 to 82 and lower than female afterwards. In general, this implies that among the observed population, males had lower life expectancy relative to the female population.

Figure 2 shows the graduated rates for male participants using  $h = 10$ ,  $h = 50$  and  $h = 100$ , respectively, for different degrees of fitness. On the other hand, Figure 3 shows the graduated rates for female participants using  $h = 10$ ,  $h = 50$  and  $h = 100$ , respectively, for different degrees of fitness.

Table 1 shows the different values of  $M$  in the Whittaker-Henderson graduation formula. Recall that Whittaker-Henderson graduation formula minimizes  $M$  to obtain the optimal values of graduated mortality rates with respect to the arbitrarily assigned  $h$  and  $z$ . The table shows that the minimum value for  $M$  was obtained by assigning  $h = 10$  and  $z = 4$ . These rates will be adjusted using mortality trend model and partial credibility theory to attain the desirable values for the mortality rates of older ages.

Table 1: Minimized  $M$  for all scenarios.

$h$	Male				Female			
	$z = 2$	$z = 3$	$z = 4$	$z = 5$	$z = 2$	$z = 3$	$z = 4$	$z = 5$
10	3.64E-05	4.99E-06	4.18E-06	7.00E-06	5.00E-05	7.08E-06	5.23E-06	1.00E-05
50	1.31E-04	1.05E-05	9.06E-06	2.29E-05	1.48E-04	2.60E-05	1.98E-05	4.38E-05
100	2.31E-04	2.10E-05	1.61E-05	4.12E-05	2.23E-04	4.95E-05	3.81E-05	8.60E-05

Figure 4 shows the evolution of the final mortality rates from the crude mortality rates for males and females as a result of the adjustments made to obtain the desired estimate. The graduated mortality rates under the assumption of  $h = 10$  and  $z = 4$  were chosen. It was adjusted using mortality trend models based on published mortality rates and partial credibility theory.

Figure 5 shows the trend of mortality rates for males and females. 2007 and 2011 rates were used to estimate mortality rates that were used as the expected mortality rates to adjust the graduated rates using partial credibility.

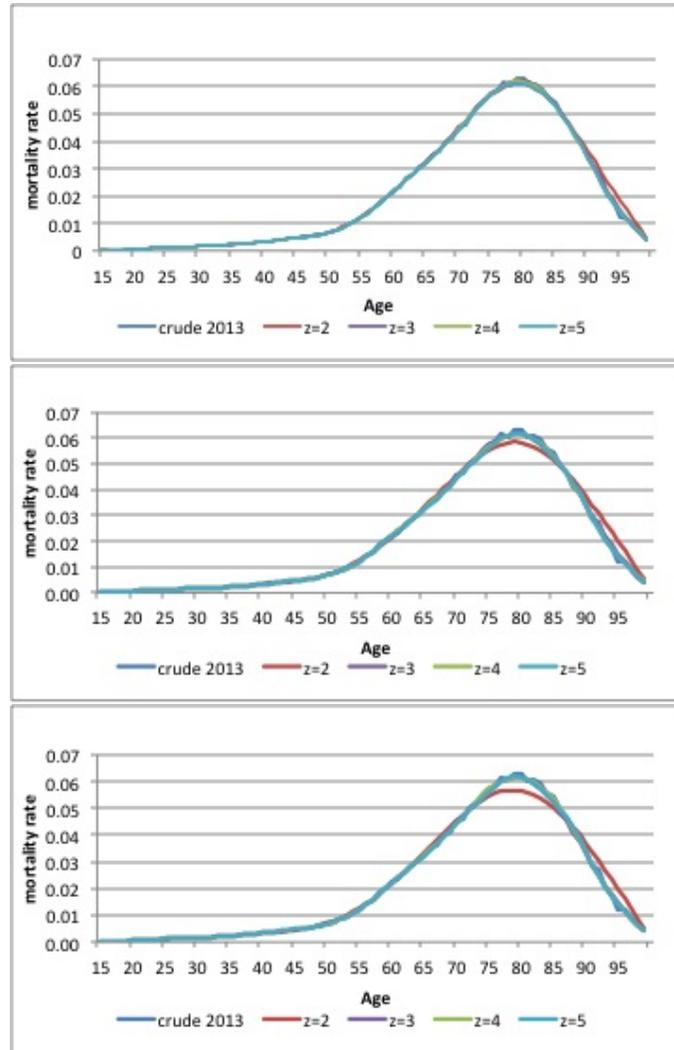


Figure 2: Graduated mortality rates of males using  $h = 10$  (top),  $h = 50$  (middle), and  $h = 100$  (bottom) for the year 2013.

Figure 6 shows the final mortality rates with the desired behavior, without deviating too far from the actual mortality rates for the younger population.

## 4 Conclusion

In this paper, mortality rates were estimated using the Whittaker-Henderson Graduation Technique, based on 10 years of Philippine Social Security System data. Among all the scenarios tested, the graduated mortality rates under the assumption of  $h = 10$  and  $z = 4$  produced the minimum value of  $M$  and produced the best estimate for the 2013 mortality rates based on the SSS members, a reasonable sample on which to base mortality rates in the Philippines. Researchers can extend this study by exploring Whittaker-Henderson

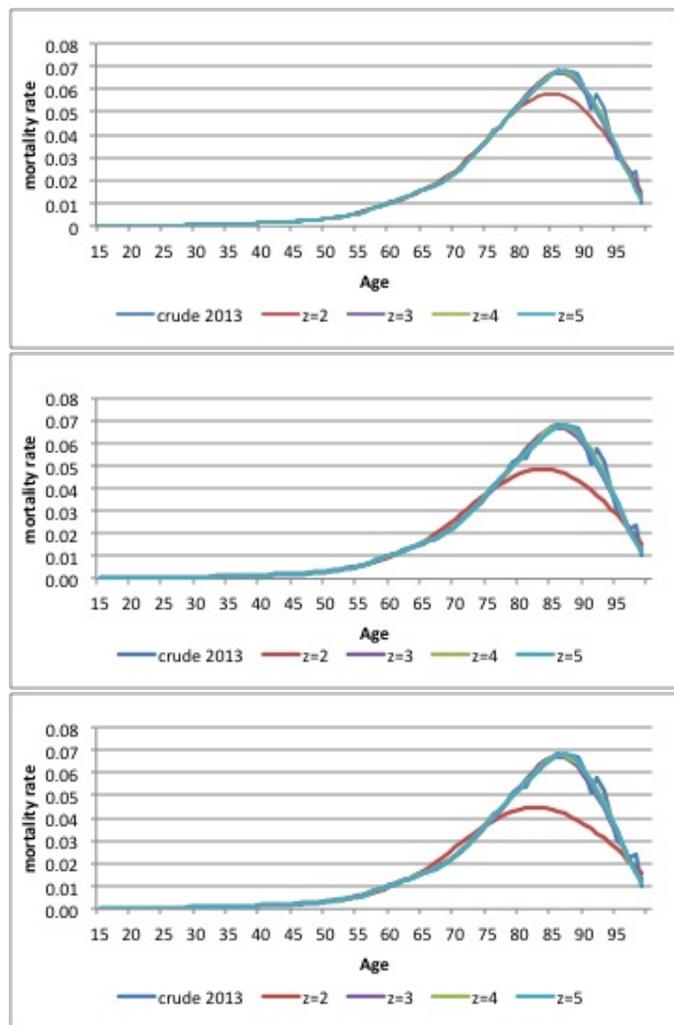


Figure 3: Graduated mortality rates of females using  $h = 10$  (top),  $h = 50$  (middle), and  $h = 100$  (bottom) for the year 2013.

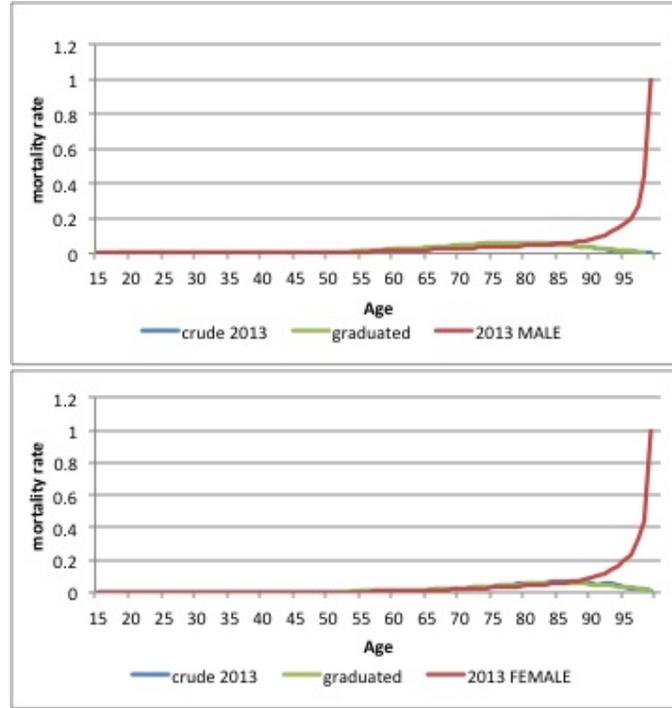


Figure 4: Crude and graduated mortality rates of males (top) and females (bottom), assuming  $h = 10$  and  $z = 4$ , for the year 2013.

graduation techniques and applying different assumptions. Researchers can also focus on the credibility testing part of this study to improve the estimation of mortality rates for older ages.

## 5 Acknowledgments

The author would like to thank SSS Actuarial Department for the permission to use confidential data necessary to make this study happen. The author would also like to acknowledge key persons from the Institute of Mathematics who motivated the author to conduct such study, particularly Daryl Saddi, Dr. Joma Escaner and Dr. Guido David.

## References

- [1] Batten, Robert W. Mortality Table Construction. Englewood Cliffs, N.J.: Prentice-Hall, 1978.
- [2] Bowers, Newton L., Gerber, Hans U., Hickman, James C., Jones, Donald A., Nesbitt, Cecil J. Actuarial Mathematics, 2nd Edition. Schaumburg, IL: Society of Actuaries, 1997.
- [3] Gujarati, Damodar N. Basic Econometrics, 4th Edition. New York, NY: McGraw-Hill, 2004.

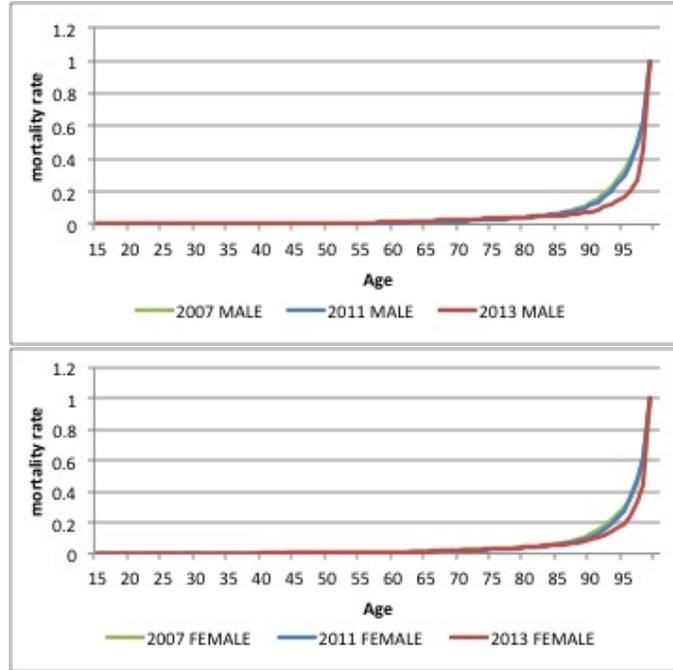


Figure 5: Mortality rate trends of males (top) and females (bottom) for the year 2013, obtained using graduated rates adjusted by partial credibility.

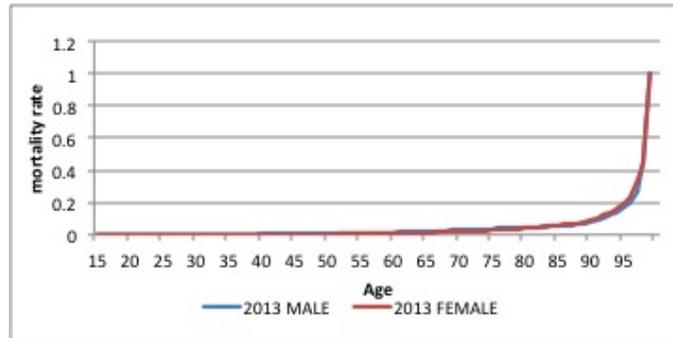


Figure 6: Final mortality rates for males (top) and females (bottom) for the year 2013.

- [4] London, Dick. *Survival Models and their estimation*, 3rd Edition. Winsted, CT: Actex Publications, 1997.
- [5] London, Dick. *Graduation: The revision of estimates*, Winsted, CT: Actex Publications, 1985.
- [6] Mahler, Howard C., Dean, Curtis G. *Credibility*, Chapter 8. In: *Casualty Actuarial Society – Foundations of Casualty Actuarial Science*, 4th Ed. Casualty Actuarial Society, 2001: 485–659.