Functional data modelling approaches for analysing and predicting trends in incidence rates – an application to falls injury

Shahid Ullah & Caroline F Finch

School of Human Movement and Sport Sciences, University of Ballarat,

VIC 3353, Australia

Address for correspondence:

Dr Shahid Ullah

School of Human Movement and Sport Sciences

University of Ballarat

Mt Helen, Victoria, 3353, Australia

Ph: +61 3 5327 9061, Fax: +61 3 5327 9478

Email: s.ullah@ballarat.edu.au

ABSTRACT

The knowledge of epidemiological information is an important to understand the behaviour of injury incidence trends over time and develop models to predict incidence patterns for the next 20 years and beyond, to inform public health and economic policy in relation to the increasing burden of injuries. Functional data model (FDM), is proposed to make long run prediction, with confidence intervals, of age-specific annual incidence of fall-related severe head injuries of people 80+ year olds in Finland during 1970-2004. The idea behind FDM is to express discrete observations in the form of a function, and then to draw information from a collection of functional data by applying concepts from multivariate data analysis. By applying it to the Finnish data, a significant increase was prediction suggesting that incidence fall-related severe head injuries in older aged 80+ years in Finland would increase severely about 2.3-2.6 folds higher in the year 2024 than it was in 2004. The predictions have far narrower confidence bands with an average of plus 609 and minus 434 injuries per 100,000 persons for all age groups, including uncertainty concerning the estimated trend. The greater predictive accuracy of FDM arises for several reasons: (a) It has approximately 55% less prediction error than traditional ordinary least square (OLS) method. (b) It produces prediction intervals for future injury incidence rates which give exact guidance as to their likely accuracy.

Keywords: incidence, functional data model, trend, prediction, injurious falls

Injuries resulting from falls in older people are a major public health concern in terms of morbidity, mortality and the associated cost of health and social services for older people as life expectancy has risen in most industrialized countries (1-8). Falls are a leading specified cause of injury death in people aged 80+ years, with mortality rates increasing rapidly with age (3,9-16). It is expected that there will be a significant increase in the size of the older population in most countries over the next 20 years, with a coincident increase in the number of falls (17-23) and this will escalate the size of the burden that injurious falls will place on both individuals and their communities (24).

Although fall-related injuries are a recognised major public health concern, until recently relatively little has been known about trends in falls injury rates (24-27). Recent increased interest in modelling falls injury rates (3,5-7,28-30) and projections (2,18,25,31-34), has largely been driven by the need for good data to inform government policy and planning in relation to fundamental changes in welfare policy needed in response to the projected population demographic shifts (10,35). Importantly, such models and predictions will only be useful in the long term if they are accurate, based on good quality data and generated through application of robust appropriate statistical methods.

Much injury epidemiology is largely descriptive in nature, and it has only been recently that sophisticated statistical modelling approaches have been

applied and developed to handle complex injury data (36,37). To date, there have been many studies on injury incidence predictions, the vast majority based on the OLS method (18,27,33,34). Poisson and negative binomial regression have also been most commonly used to model and predict injury incidence in recent years (38). Common problems associated with regression model predictions are that there are strong parametric assumptions underlying the statistical models such as that trend in incidence rates and the resources used to treat these injuries will remain unchanged (39-41). Age-period-cohort methods have also been used for injury incidence predictions with Poisson regression modelling used to quantify the age, period, and cohort effects (42,43). However, there is ongoing debate about the most appropriate models for this adjustment (44). The problem arises with the sensitivity of the projections to the most recent changes in cohort effects (39,45,46).

The introduction of stochastic methods for predicting rates in other health areas has been shown to have significant advantages for better understanding trends, risk factor relationships, and the effectiveness of preventive measures (39,47-49). A major advantage of stochastic methods is that they can measure prediction uncertainty through the estimation of prediction intervals of future injury incidence. An important milestone has been the methodology proposed by Lee et al. (48,50) for the modelling and extrapolating of long-term trends in mortality rates used to predict US mortality rates to 2065. More recently, functional data methods (FDM), have

started to receive attention, particularly in demographic and medical applications (39,47,51,52). To our knowledge, FDM have not been widely applied to epidemiological studies for the prediction of incidence rates, including studies relating to injury incidence associated with falls in older people.

Using the specific example of fall-related head injuries in older people, this paper shows how to apply FDM to describe time trends of age-specific incidence rates and their future projections. The modelled incidence rates are discrete observations for specified ages and calendar year and predictions are made into the future. The essence of the FDM is to express these discrete values in the form of a function, and then to draw statistical information from the resultant collection of functional data by applying multivariate data analysis approaches. This approach has the potential to improve the likely accuracy of these predictions, compared to current methods in the literature such as the more common OLS models. In applying FDM, there are a number of key statistical issues that need to be addressed and this paper provides guidance and examples for: useful graphical tools for effective data representation; making no parametric assumptions about the function or the observed data; ensuring the functions are smooth to enable modelling of the data's dynamic behaviour; handling complexity and interdependencies between variables; and deriving prediction intervals to assess the likely accuracy of projections.

In the specific application area of modelling fall-related head injury this study a) models age-related changes in fall-related head injury incidence time trends using FDM; b) predicts future age-related changes in these incidence rates using a recently developed prediction method (47); and c) measures the accuracy of predictions for future fall related head injuries.

METHODS

The Data

Both fatal and non-fatal head injuries can be a serious consequence of falls in older people. A recent study in Finland showed that the secular trend in the incidence of these injuries in people aged 80+ years is rising (32). That study also compared actual, observed fall-related head injury incidence rates to previous predictions of those rates by the same authors (18). The earlier study adopted the common OLS method for projecting long term trends from 1970 to 1995. The presence of published predictions (18) and data on actual observations years later (32) prompted this comparison of two predictive approaches (OLS and FDM) for the incidence of fall-related head injuries in Finland.

The actual data underpinning the published analysis (32) was obtained directly from the first author of that study (Pekka Kannus, UKK Institute for Health Promotion Research, Tampere, Finland). The data originated from the Finnish National Hospital Discharge Register and cases were identified

through relevant *International Classification of Diseases* (ICD) primary and secondary diagnoses and E-code causes of injury (E-code). Full details of the case selection is given in the original Finnish study (32). A severe fall-related head injury was defined as a head injury as a consequence of a fall from standing height of $\leq 1\text{m}$ and that resulted in hospitalization. Data obtained was the annual number and incidence of fall-related severe head injuries during the period 1970-2004, in people aged 80-84, 85-89, and 90+ years. All incidence rates were age-adjusted by direct standardization using the mean population of persons aged 80+ years between 1970 and 2004 as the standard population. The age-specific injury incidence was calculated in 5-year age groups and was expressed as the number of cases per 100,000 persons per year. A more detailed description of the data has been published elsewhere by Kannus et al. (18,32).

The Functional Data Model

Smoothing

Within each age group, the annual fall-related severe head injury incidence rates were considered as a function of age, centred around the mid-point of the age group. The age-specific injury incidence rates for each year were plotted and the incidence rates defined as age-incidence curves. The age-specific injury incidence rates for each year were smoothed using penalized regression splines (47) to estimate the age-incidence curves. Since the incidence in the Finnish data increased rapidly with age, the

incidence curves were assumed to increase monotonically with age. To capture the trends in incidence, constrained penalized regression splines (53,54) with a monotonic constraint applied to represent the smooth curves were used according to the method of (47). The `gam()` function in R from the *mgcv* package was used with the penalized regression splines to represent the curves. The use of penalized regression splines has a number of advantages in that a) the smoothness conditions can easily be adopted to the nature of the incidence data analysed thereby reducing the noise in the incidence curves; and b) the underlying process generating the age-incidence curves is then continuous and smooth. This continuity property is particularly relevant because it leads to functional data that can be analyzed with FDMs (55-57).

Functional principal component

The smoothed curves (or functional observations) were then modelled by FDM to estimate the functions representing the age-incidence relationship, i.e the basis functions, according to the method of Hyndman and Ullah (47). Appendix 1 gives more details of the functional data approach. As all parameters on the right hand side of Equation 2 in Appendix 1 are unobservable, fitting the model using the ordinary least squares method is impossible. To overcome the situation, functional principal components (FPC) decomposition was applied to the smoothed curves to estimate the basis functions which minimise the mean integrated squared error (MISE)

(47) [see Appendix 1 for details]. The basis functions show the shape of the variability of the incidence curves. The extracted functions can be easily interpreted because the basis functions are defined in the same domain as the original incidence curves. The predictions of each incidence curve are sensitive to the number of basis functions. A higher number of basis functions does not necessarily give better forecasts or improve the prediction accuracy (51). Although, there are many ways for selecting the basis functions, Hyndman and Ullah (47) proposed a procedure which minimizes the MISE.

Prediction

The functional principal components method described in the previous section also estimated the time series coefficients of the model. The values of coefficients form a time series, with one value for each year of data. Standard time series methods can be used to model and predict these time series coefficients. The random walk with drift is almost exclusively used for predicting rates in other health areas (48,58). However, the exponential smoothing state space model is found to be statistically adequate for the current dataset and we therefore consider this model as the driving process. Analytical expressions (averages of multi-step-ahead predictions and the corresponding standard deviation of the prediction errors) were used to construct the prediction intervals (59,60). Appendix 2 gives the detailed analytical expressions.

Prediction accuracy

The integrated squared prediction error (ISPE) was used to evaluate the accuracy of the estimated predictions of future incidence rates [see Appendix 2 for details]. In designing the accuracy measures of future age-incidence rates, an out-of-sample test is performed (61). An out-of-sample evaluation of prediction accuracy begins with the division of the data set into a fit period and a test period. The fit period is used to identify and estimate the model while the test period is reserved to assess the model's prediction accuracy. Based on the fitting period 1970-1994, the FDM predictions of injury incidence rates for 1995-2004 were directly compared with the published actual rates for 1995-2004 through averaging of the ISPE (47) .

All statistical analyses were performed in R version 2.8.1 (62), operating on a Windows XP professional platform.

RESULTS

Figure 1a shows the trends in age-specific incidence rates for head injuries in Finland over the period 1970-2004. Incidence rates were increased with age and time. As described by Kannus and colleagues (32), over the 35 year period, the relative increases in severe fall-related head injury incidence rates were 159% for 80-84 year olds, 134% for 84-89 year olds and 132% for the 90+ year olds. There was a large increase in

incidence rates over the last decade in all age groups, though this was most marked in the 90+ year age group. The characteristics of these age-group specific incidence rates clearly indicate the appropriateness of the FDM approach for making predictions based on it. Firstly, there is a high degree of variability of incidence rates at older ages and secondly, the general increase in incidence rates over time is not uniform across all age-groups.

The five-year age group incidence rates were converted into functional data by estimating a smooth curve through the observations, taking the centre of each age group as the point of interpolation. A closer look at a comparatively stable data shown in Figure 1b, suggests that the incidence rates vary fairly smoothly and regularly within each year.

<Insert Figure 1 about here>

Following the procedures outlined in Appendices 1 and 2, the model diagnostics suggested that the first three basis functions (age-incidence relationships) were sufficient to describe the injury incidence curves. These three functions explained 97.5%, 2.0% and 0.5% of the total variation, respectively, and can therefore be taken as being jointly adequate for describing the incidence data.

The average incidence curve is shown in Figure 2a. As expected, the average incidence rates sharply increase with age. The first basis function

and the first coefficient are shown in Figures 2b and c. Because the second and third basis functions collectively explained only 2.5% of the total variation, they are less important than the first one. The first coefficient indicates a fairly steady increase in fall-related severe head injuries over time, while the basis function indicates that the increase has been faster in 90+ year olds.

<Insert Figure 2 about here>

Figure 3 shows the predicted age-specific fall-related severe head injury incidence rates for 2005-2024 with 95% prediction intervals, together with the observed rates for 1970-2004. Combining the predicted coefficients with the estimated basis functions yields predicted age-specific incidence rates. The data are averaged on 5-year age range from 80 to 94 years. The predicted incidence rates are estimated to increase sharply over time, as was also shown by Kannus et al (32). The predicted incidence rates for 2024 are 1108 injuries per 100,000 persons with a 95% confidence interval of 763 to 1611 (80-84 year olds), 1573 injuries per 100,000 persons with an interval of 1156 to 2142 (85-89 year olds) and, 1916 injuries per 100,000 persons with an interval of 1375 to 2672 (90+ year olds), representing 2.3-2.6 fold increases over the 2004 rates for all age groups.

<Insert Figure 3 about here>

Comparison of the FDM versus OLS predictions

The results from FDM predictions were compared with the published predictions based (32) on the OLS method. The methods were applied to the Finish injury incidence rates for years 1970-1994 and predict incidence for years 1995-2004. Based on the ISFE approach of Appendix 2, direct comparison of the FDM and OLS predictions and the actual rates are shown in Figure 4. The 95% prediction intervals for FDM predictions are also shown. The FDM out-of-sample prediction performed better than OLS method. The OLS predictions for longer time period were beyond the prediction intervals of FDM method in all age groups. Although there were little differences for the first prediction year, thereafter the FDM predictions were closest to the actual values. The advantage of the FDM approach becomes apparent in that it has approximately 55% less prediction error than the OLS method.

<Insert Figure 4 about here>

DISCUSSION AND CONCLUSION

Functional Data Methods provide a relatively novel modelling and prediction approach to estimating incidence trends. However, to date their application to epidemiological data has been limited and, to our knowledge, non-existent in the injury literature. This paper therefore represents the first

application of FDM to a significant injury epidemiology problem, from any context.

With ageing populations worldwide, and the expected increasing burden of fall-related morbidity and mortality in older people, it is becoming more and more important to understand the likely trends in all types of serious fall-related injury incidence trends over time. Moreover, policy makers need information about predicted future injury rates to inform their decision making about public health and economic investments to reduce the burden of these injuries well into the future. It is critical that such predictions are robust and based on the best available statistical modelling approaches so as to minimise possible error in the forecasts. This is also true for other areas of public health and injury prevention.

The FDM is a natural extension of methods developed for mortality and fertility prediction that have evolved over the last two decades (47,48,50). The methodology has since become very widely used in demographic applications and there have been various extensions and modifications proposed (50,52,63-65). Somewhat surprisingly, the use of FDM in epidemiological applications has not been widespread.

The FDM is effective for exploring complex multivariate functional relationships and it has the major strength of being able to model the functional form of age-related changes in incidence rates over time and to

make predictions for specific age groups (47). In the injury context, this facilitates understanding of the different patterns of injury rate progression in different age groups. By applying it to the Finnish data, we have shown that it can provide more precise estimates of trends in fall-related severe head injuries among older people than is possible with the more common OLS method projections.

A second major strength of the FDM approach is the improved modelling of inconsistent increases in incidence rates over time, particularly the high degree of variability of incidence rates at older ages. The OLS approach with auto-correlated errors (18,32) was unable to capture the unstable trend and high variability of injury incidence rates at older ages. While it is important to understand the reason for such inconsistencies, it is equally important that statistical models and the predictions they generate are able to factor them into their estimation processes. Because FDM has not been previously applied to injury data, it is possible that failure to account for this could have lead to incorrect or non-optimal epidemiological inferences about future injury outcomes. The FDM approach of initially smoothing the data and then using these smoothed observations for the modelling and prediction estimation, is a major methodological improvement over methods that simply fit linear/nonlinear trends to observed incidence data, such as OLS methods.

Finally, FDM models provide prediction intervals for future incidence projections. All previous injury prediction studies have only presented point estimates (2,18,25,31,32,66). Although the calculation of prediction intervals is possible using OLS methods, they are likely to be inaccurate outside the range of injury incidence because it is unlikely that the linear relationship assumption would hold for future incidence (67). The methods presented in this paper could easily be incorporated into other public health and injury incidence applications, including the modelling of known risk factors and economic evaluations. As the shape of the incidence curves varies with risk factors over time, the FDM approach enables the models and predictions to pick up subtle variations for these factors (39), such as the age effects in the data analysed in this paper.

In the context of application of FDM to the incidence of fall-related severe head injuries in Finns aged 80+ years, this study provides an improved understanding of current trends (particularly in terms of the age relationships). Moreover, this paper clearly demonstrates the greater predictive accuracy of FDM over OLS methods, the former having 55% less prediction error for the same data, and the generation of prediction intervals that give information about the likely accuracy of predictions.

In summary, this paper describes FDM methods and their important application to epidemiology. In doing so, it establishes FDM as a preferred

modelling approach for incidence trends and predictions, with the potential for many significant applications across a range of epidemiological issues.

ACKNOWLEDGEMENT

Dr Shahid Ullah was supported by an Injury Trauma and Rehabilitation (ITR) Research Fellowship funded through a National Health and Medical Research Council (NHMRC) Capacity Building Grant in Population Health. Prof Caroline Finch was supported by an NHMRC Principal Research Fellowship. The authors wish to thank Professor Pekka Kannus, Chief Physician and Head, Injury & Osteoporosis Research Centre, UKK Institute for Health Promotion Research, Tampere, Finland for providing the data analysed in this study.

REFERENCES

- (1) Masud T, Morris RO. Epidemiology of falls. *Age Ageing*. 2001; 30 (Suppl 4):3-7.
- (2) Kannus P, Niemi S, Palvanen M, et al. Continuously increasing number and incidence of fall-induced, fracture-associated, spinal cord injuries in elderly persons. *Arch Intern Med*. 2000;160(14):2145-2149.
- (3) Kannus P, Parkkari J, Koskinen S, et al. Fall-induced injuries and deaths among older adults. *JAMA*. 1999; 281(20):1895-1899.
- (4) Tinetti M, Speechley M. Prevention of falls among the elderly. *N Engl J Med*. 1989;320(16):1055-1059.
- (5) Boufous S, Finch C, Close J, et al. Hospital admissions following presentations to emergency departments for a fracture in older people. *Inj Prev*. 2007;13(3):211-214.
- (6) Boufous S, Finch C, Lord S, et al. The increasing burden of pelvic fractures in older people, New South Wales, Australia. *Injury*. 2005; 36(11):1323-1329.
- (7) Boufous S, Finch CF, Lord SR. Incidence of hip fracture in New South Wales: are our efforts having an effect? *Med J Aust*. 2004;180:623-626.
- (8) Centers for Disease Control and Prevention. *Web-based Injury Statistics Query and Reporting System (WISQARS)* [online]. NCIPC, CDC (producer). [accessed 2 Jan 2008]. Available from: URL: Available at: www.cdc.gov/ncipc/wisqars.
- (9) Australian Institute of Health and Welfare. *Australia's Health 2008: Section 5 Diseases and Injuries*. Australian Institute of Health and Welfare, Canberra 2008;Cat. no. AUS 99.
- (10) Moller J. Current costing models: are they suitable for allocating health resources? The example of fall injury prevention in Australia. *Accid Anal Prev*. 2005;37(1):25-33.
- (11) Hindmarsh DM, Hayen A, Finch CF, et al. Relative survival after hospitalisation for hip fracture in older people in New South Wales, Australia. *Osteoporos Int*. 2009;20(2):221-229.
- (12) Kannus P, Parkkari J, Niemi S, et al. Fall-induced deaths among elderly people. *Am J Public Health*. 2005;95(3):422-424.

- (13) Orces CH. Trends in fall-related mortality among older adults in Texas. *Tex Med*. 2008;104(5):55-59.
- (14) Scott VJ, Gallagher EM. Mortality and morbidity related to injuries from falls in British Columbia. *Can J Public Health*. 1999;90(5):343-347.
- (15) Stevens JA, Dellinger AM. Motor vehicle and fall related deaths among older Americans 1990-98: sex, race, and ethnic disparities. *Inj Prev*. 2002; 8(4):272-275.
- (16) Sattin RW. Falls among older persons: a public health perspective. *Annu Rev Public Health*. 1992;13:489-508.
- (17) Stevens JA, Adekoya N. Brain injury resulting from falls among elderly persons. *JAMA*. 2001;286(21):2665-2666.
- (18) Kannus P, Palvanen M, Niemi S, et al. Increasing number and incidence of fall-induced severe head injuries in older adults: nationwide statistics in Finland in 1970-1995 and prediction for the future. *Am J Epidemiol*. 1999;149(2):143-150.
- (19) Rubenstein LZ, Josephson KR. The epidemiology of falls and syncope. *Clin Geriatr Med*. 2002;18(2):141-158.
- (20) Fleming J, Matthews FE, Brayne C, et al. Falls in advanced old age: recalled falls and prospective follow-up of over-90-year-olds in the Cambridge City over-75s cohort study. *BMC Geriatr*. 2008;8:6.
- (21) Chang JT, Morton SC, Rubenstein LZ, et al. Interventions for the prevention of falls in older adults: systematic review and meta-analysis of randomised clinical trials. *Br Med J*. 2004;328(7441):680.
- (22) Stevens JA, Mack KA, Paulozzi LJ, et al. Self-reported falls and fall-related injuries among persons aged >65 years - United States, 2006. *J.Safety Res*. 2008;39(3):345-349.
- (23) Peel NM, Kassulke DJ, McClure RJ. Population based study of hospitalised fall related injuries in older people. *Inj Prev* 2002;8(4):280-283.
- (24) Mann E, Icks A, Haastert B, et al. Hip fracture incidence in the elderly in Austria: an epidemiological study covering the years 1994 to 2006. *BMC Geriatr*. 2008;8:35.
- (25) Kannus P, Niemi S, Palvanen M, et al. Rising incidence of fall-induced injuries among elderly adults. *J Public Health*. 2005;13(4):212-215.

- (26) Gannon B, O'Shea E, Hudson E. Economic consequences of falls and fractures among older people. *Ir Med J*. 2008;101(6):170-173.
- (27) Kannus P, Palvanen M, Niemi S, et al. Rate of proximal humeral fractures in older Finnish women between 1970 and 2007. *Bone*. 2009; 20: 43-46.
- (28) Robertson MC, Campbell AJ, Herbison P. Statistical analysis of efficacy in falls prevention trials. *J Gerontol A Biol Sci Med Sci* 2005;60:530-534.
- (29) Lord SR. Visual risk factors for falls in older people. *Age Ageing*. 2006 35(Suppl 2):42-45.
- (30) Boufous S, Finch C, Lord S, et al. The epidemiology of hospitalised wrist fractures in older people, New South Wales, Australia. *Bone*. 2006; 39(5):1144-1148.
- (31) Kannus P, Palvanen M, Niemi S, et al. Alarming rise in the number and incidence of fall-induced cervical spine injuries among older adults. *J Gerontol A Biol Sci Med Sci*. 2007;62(2):180-183.
- (32) Kannus P, Niemi S, Parkkari J, et al. Alarming rise in fall-induced severe head injuries among elderly people. *Injury* 2007; 38(1):81-83.
- (33) Kannus P, Niemi S, Parkkari J, et al. Hip fractures in Finland between 1970 and 1997 and predictions for the future. *Lancet*. 1999; 353(9155): 802-805.
- (34) Kannus P, Palvanen M, Niemi S, et al. Increasing number and incidence of low-trauma ankle fractures in elderly people: Finnish statistics during 1970-2000 and projections for the future. *Bone*. 2002;31(3):430-433.
- (35) Finch CF, Hayen A. Governmental health agencies need to assume leadership in injury prevention. *Inj Prev* 2006;12(1):2-3.
- (36) MacKenzie EJ. Epidemiology of injuries: current trends and future challenges. *Epidemiol Rev*. 2000;22(1):112-119.
- (37) Bangdiwala SI. Random or accidental - what is the goodness of the fit? *Int J Inj Contr Saf Promot*. 2009;16(1):59-60.
- (38) Rivara FP, Cummings P, Koepsell TD, et al. *Injury Control: A Guide to Research and Program Evaluation*. Cambridge University Press, New York; 2001.

- (39) Erbas B, Hyndman RJ, Gertig DM. Forecasting age-specific breast cancer mortality using functional data models. *Stat Med*. 2007;26(2):458-470.
- (40) Mathers CD, Loncar D. Projections of global mortality and burden of disease from 2002 to 2030. *PLoS Med*. 2006;3(11):2011-2030.
- (41) Moller J. Changing health resource demands for injury due to falls in an ageing population. *N S W Public Health Bull*. 2002;13(1-2):3-6.
- (42) Li G, Shahpar C, Grabowski JG, Baker SP. Secular trends of motor vehicle mortality in the United States, 1910–1994. *Accid Anal Prev*. 2001; 33(3):423-432.
- (43) Wildner M, Casper W, Bergmann KE. A secular trend in hip fracture incidence in East Germany. *Osteoporos Int*. 1999;9(2):144-150.
- (44) Icks A, Haastert B, Wildner M, et al. Trend of hip fracture incidence in Germany 1995-2004: a population-based study. *Osteoporos Int*. 2008; 19(8):1139-1145.
- (45) Holford TR. Understanding the effects of age, period, and cohort on incidence and mortality rates. *Annu Rev Public Health*. 1991;12(1):425-457.
- (46) Willets RC. The cohort effect: Insights and explanations. *Br Actur J*. 2004;10:833-877.
- (47) Hyndman RJ, Ullah S. Robust forecasting of mortality and fertility rates: A functional data approach. *Comput Stat Data Anal*. 2007;51(10):4942-4956.
- (48) Lee RD, Carter LR. Modeling and forecasting U.S. mortality. *J Am Stat Assoc*. 1992;87:659-675.
- (49) Lee R, Miller T. An approach to forecasting health expenditures, with application to the U.S. Medicare system. *Health Serv Res*. 2002; 37(5):1365-1386.
- (50) Lee RD, Miller T. Evaluating the performance of the Lee-Carter method for forecasting mortality. *Demography*. 2001;38(4):537-549.
- (51) Hyndman RJ, Booth H. Stochastic population forecasts using functional data models for mortality, fertility and migration. *Int J Forecast*. 2008 24(3): 323-342.

- (52) De Jong P, Tickle L. Extending Lee-Carter mortality forecasting. *Math Popul Stud.* 2006;13(1):1-18.
- (53) Ruppert D, Wand MP, Carroll RJ. *Semiparametric Regression*. New York: Cambridge University Press; 2003.
- (54) Simonoff JS. *Smoothing Methods in Statistics*. New York: Springer-Verlag; 1996.
- (55) Ramsay JO, Dalzell CJ. Some tools for functional data analysis. *J R Stat Soc Series B Stat Methodol.* 1991;53(3):539-572.
- (56) Ramsay JO, Silverman BW. *Applied Functional Data Analysis*. New York: Springer-Verlag; 2002.
- (57) Ramsay JO, Silverman BW. *Functional Data Analysis. 2nd ed.* New York: Springer-Verlag; 2005.
- (58) Lee RD, Tuljapurkar S. Stochastic population forecasts for the United States: Beyond high, medium, and low. *J Am Stat Assoc.* 1994; 89(428): 1175-1189.
- (59) Hyndman RJ, Koehler AB, Ord JK, Snyder RD. Prediction intervals for exponential smoothing using two new classes of state space models. *J Forecast.* 2005;24(1):17-37.
- (60) Hyndman RJ, Koehler AB, Snyder RD, et al. A state space framework for automatic forecasting using exponential smoothing methods. *Int J Forecast.* 2002;18(3):439-454.
- (61) Tashman LJ. Out-of-sample tests of forecasting accuracy: an analysis and review. *Int J Forecast.* 2000;16(4):437-450.
- (62) R Development Core Team. A Language and Environment for Statistical Computing. R Foundation for Statistical Computing: Vienna, Austria. 2004; Available at: <http://www.R-project.org>.
- (63) Renshaw A, Haberman S. Lee-Carter mortality forecasting: A parallel generalized linear modelling approach for England and Wales mortality projections. *Appl Stat.* 2003; 52(1):119-137.
- (64) Lee RD. Modeling and forecasting the time series of US fertility: Age distribution, range, and ultimate level. *Int J Forecast.* 1993; 9(2):187-202.
- (65) Booth H, Maindonald J, Smith L. Applying Lee-Carter under conditions of variable mortality decline. *Popul Stud.* 2002; 56(3):325-336.

(66) O'Connor PJ. Forecasting of spinal cord injury annual case numbers in Australia. *Arch Phys Med Rehabil.* 2005; 86(1): 48-51.

(67) Makridakis S, Wheelwright SC, Hyndman RJ. *Forecasting: Methods and applications.* 3rd ed. Wiley, New York; 1998.

FIGURE LEGENDS

Figure 1. Age-specific incidence rates of fall-induced severe head injuries in people aged 80+ years and older in Finland, 1970-2004. (a) Actual incidence rates (■ = 80-84 years, ● = 85-89 years, ▲ = 90+ years) and (b) Smoothed incidence rates

Figure 2. Principal components from the functional data models (FDM) for incidence of fall-induced severe head injuries in Finland. A decomposition of order $K = 3$ has been used. (a) the mean of the incidence rates, (b) the first basis function and (c) the coefficient associated with basis function.

Figure 3. Predicted age-specific incidence of fall-induced severe head injuries in people of age 80 years and older in Finland for the years 2005-2024. 95% prediction intervals are also in the figure. (a) 80-84 years, (b) 85-89 years and (c) 90+ years.

Figure 4. Prediction accuracy of FDM and OLS methods of fall-induced severe head injuries. The fitting period is 1970-1994 and the prediction period is 1995-2004. The FDM 95% prediction intervals are also in the figure. (a) 80-84 years, (b) 85-89 years and (c) 90+ years.

Appendix 1: Functional data model (FDM) and mean integrated squared error (MISE)

Let $y_t(x)$ denote the observed fall-related head injury incidence for age x in year t . Then we assume the following model

$$y_t(x) = f_t(x) + \sigma_t(x)\varepsilon_t(x) \quad [1]$$

$$f_t(x) = \mu(x) + \sum_{k=1}^K \beta_{t,k} \phi_k(x) + e_t(x) \quad [2]$$

where $f_t(x)$ is an underlying head injury incidence smooth function of x observed with error, $\varepsilon_t(x)$ are independently and identically distributed standard normal variates and $\sigma_t(x)$ allows the variance to change with age and year according to the nature of the data. To smooth the incidence data, penalized regression splines are used.

The second equation (2) describes the dynamics of $f_t(x)$ evolving through time. In this equation, $\mu(x)$ is the mean of smooth incidence curves $f_t(x)$ across years and $e_t(x)$ is the model error. The age component $\phi_k(x)$ is a set of orthogonal basis functions or principal components which modifies the main time trend according to whether change at a particular age is faster or slower than the main trend (and in the same or opposite direction). The model assumes that $\hat{\phi}_k(x)$ is invariant over time. The time

component, $\beta_{t,k}$, are time series coefficients which capture the overall time trend in $f_t(x)$ at all ages. The model makes no assumptions about the functional form of the trend in $\beta_{t,k}$.

Let $e_t(x) = f_t(x) - \hat{f}_t(x)$ denote the fitted error from the above equation. For a given number of basis functions, K , The functional principal component decomposition, applied to the smooth incidence curves $f_t(x)$, gives the basis functions $\phi_k(x)$ which minimize the mean integrated squared error

$$MISE = \frac{1}{n} \sum_{t=1}^n \int e_t^2(x) dx, \quad [3]$$

Appendix 2: Prediction intervals and integrated squared prediction error (ISPE)

Suppose we have injury incidence data up to year $t = n$, and we wish to estimate future incidence rates to year $n+h$ $y_t(x)$ for $t = n+1, \dots, n+h$ and all x . Let $\hat{\beta}_{n,k,h}$ denote the h -step ahead forecast of $\beta_{n+h,k}$. Let $\hat{y}_{n,h}(x)$ denote the h -step ahead forecast of $y_{n+h}(x)$ and let $\hat{f}_{n,h}(x)$ denote the h -step ahead forecast of $f_{n+h}(x)$. Then

$$\hat{y}_{n,h}(x) = \hat{f}_{n,h}(x) = \hat{\mu}(x) + \sum_{k=1}^K \hat{\beta}_{n,k,h} \hat{\phi}_k(x) \quad [1]$$

Following (47), we can give the following expression for forecast variance

$$V_{n,h}(x) = \text{Var}[y_{n+h}(x) | I, \phi] = \hat{\sigma}_{\mu}^2(x) + \sum_{k=1}^K u_{n+h,k} \phi_k^2(x) + v(x) + \sigma_t^2(x) \quad [2]$$

where $I = [y_t(x); t = 1, \dots, n]$ denotes all observed data, $u_{n+h,k} = \text{Var}[\beta_{n+h,k} | \beta_{1,k}, \dots, \beta_{n,k}]$ can be obtained from the time series model, $\hat{\sigma}_{\mu}^2(x)$ (the variance of the smooth estimate $\hat{\mu}(x)$) can be obtained from the smoothing method used and $v(x)$ is estimated by averaging $\hat{e}_t^2(x)$ for each x . So the 95% prediction interval for $y_t(x)$ is constructed as

$$\hat{y}_{n,h}(x) \pm z_{\alpha} \sqrt{V_{n,h}(x)} \quad [3]$$

where z_{α} is the standard normal quantile.

The accuracy of the injury incidence prediction is evaluated by computing the integrated squared prediction error. Let

$$e_{n,h}(x) = y_{n+h}(x) - \hat{y}_{n,h}(x) \quad [4]$$

denote the prediction error. Then the integrated squared prediction error is defined as

$$ISPE_n(h) = \int_x e_{n,h}^2(x) dx$$