

Journal of Visualized Experiments

Forecasting Hepatocellular Carcinoma Mortality Using a Weighted Regression Model to Estimate Cohort Effects in Taiwan

--Manuscript Draft--

Article Type:	Invited Methods Collection - JoVE Produced Video
Manuscript Number:	JoVE62253R4
Full Title:	Forecasting Hepatocellular Carcinoma Mortality Using a Weighted Regression Model to Estimate Cohort Effects in Taiwan
Corresponding Author:	I-Shiang Tzeng Taipei Tzu Chi Hospital New Taipei, Taiwan TAIWAN
Corresponding Author's Institution:	Taipei Tzu Chi Hospital
Corresponding Author E-Mail:	istzeng@gmail.com
Order of Authors:	I-Shiang Tzeng Chan-Yen Kuo Chia-Chi Wang
Additional Information:	
Question	Response
Please indicate whether this article will be Standard Access or Open Access.	Standard Access (US\$2,400)
Please specify the section of the submitted manuscript.	Cancer Research
Please indicate the city, state/province, and country where this article will be filmed . Please do not use abbreviations.	New Taipei, Taiwan
Please confirm that you have read and agree to the terms and conditions of the author license agreement that applies below:	I agree to the Author License Agreement
Please provide any comments to the journal here.	
Please indicate whether this article will be Standard Access or Open Access.	Standard Access (\$1400)
Please confirm that you have read and agree to the terms and conditions of the video release that applies below:	I agree to the Video Release

TITLE:

Forecasting Hepatocellular Carcinoma Mortality Using a Weighted Regression Model to Estimate Cohort Effects in Taiwan

AUTHORS AND AFFILIATIONS:

I-Shiang Tzeng^{1,2,3*}, Chan-Yen Kuo¹, Chia-Chi Wang⁴

¹Department of Research, Taipei Tzu Chi Hospital, Buddhist Tzu Chi Medical Foundation, New Taipei City, Taiwan

²Department of Statistics, National Taipei University, Taipei, Taiwan

³Department of Applied Mathematics; Department of Exercise and Health Promotion, Chinese Culture University, Taipei, Taiwan

⁴Department of Gastroenterology, Taipei Tzu Chi Hospital, Buddhist Tzu Chi Medical Foundation and School of Medicine, Tzu Chi University, Hualien, Taiwan

*Corresponding Author:

Dr. I-Shiang Tzeng

istzeng@gmail.com

KEYWORDS:

Hepatocellular carcinoma, age-period-cohort; multiphase method; weighted regression model

SUMMARY:

We depict a multistage method to measure a cohort effect with age data, thereby allowing data to be eliminated in many situations without sacrificing data quality. The protocol demonstrates the strategy and provides a weighted regression model for analyzing the hepatocellular carcinoma data.

ABSTRACT:

To eliminate the influence of age and period in age cycle contingency table data, a multistage method was adopted to evaluate the cohort effect. The most general primary malignant tumor of the liver is hepatocellular carcinoma (HCC). HCC is associated with liver cirrhosis with alcohol and viral etiologies. In epidemiology, long-term trends in HCC mortality were delineated (or forecasted) by using an age-period-cohort (APC) model. The HCC deaths were determined for each cohort with its weighted influence. The confidence interval (CI) of the weighted mean is fairly narrow (compared to the equally weighted estimates). Due to the fairly narrow CI with less uncertainty, the weighted mean estimation was used as a means for forecasting. With the multistage method, it is recommended to use weighted mean estimation based on a regression model to evaluate the cohort effect in the age-period contingency table data.

INTRODUCTION:

The most common primary malignant tumor of the liver is hepatocellular carcinoma (HCC). Its mortality rate ranks fifth in men and eighth in women (6% of men and 3% of women) ¹ among all malignant tumors worldwide. In Taiwan, it is the most common cancer in men and the

second most common cancer in women (21.8% of men and 14.2% of women)². It is estimated that since 2000, the annual number of HCCs diagnosed worldwide is 564,000, among which 398,000 are men and 166,000 are women³. In epidemiology, the most common way to explain the relationship between age, period, and cohort (APC) variables is that age and period influence each other to create a unique generational experience for the disease trend investigated.

Even though this conceptualization still has a precise linear connection of age + cohort = period, exposure (predictor) is not an inherent factor in a birth cohort. Instead, we propose that when changes cause different distributions of disease, there is a cohort effect. Nevertheless, since age + cohort = period, these three variables are linearly related; only if other restrictions are enforced is it impossible to generate an estimated age-period-cohort (APC) model using the linear effects of age, period, and cohort. In this study, we clarified this problem and the potential restrictions we imposed in our previous publications⁴⁻⁷.

With the slightest conjectures about the contingency table data, the multistage method⁸ provides three stages to evaluate the cohort effect. In addition, since median polish does not depend on a specific distribution or framework, it was used for various types of data, such as ratios, logarithmic ratios, and counts. Median polishing is the prime technique used in the multiphase method.

Data from a two-way contingency table⁹ were used to generate the development of the polished median. The median polishing procedure is used to eliminate the cumulative effects of age (i.e., row) and period (i.e., column) by iteratively subtracting the median from each row and each column. This procedure is often used in epidemiological data analysis¹⁰. One advantage of this technique is that no assumptions about the distribution or structure of the data in the bidirectional contingency table are required. Therefore, this technique was broadly utilized for any type of data contained in the table, such as suicide data¹¹. The APC model has also been used to describe the long-term trends of disease incidence or mortality⁵. APC models often assume that age, period, and cohort have additive effects on the logarithmic transformation of disease/mortality. To evaluate cohort effects, the described protocol generates an APC model for complete hepatocellular carcinoma (HCC) mortality analysis with weighted regression, thereby supporting reliable predictions and moderate assessment of treatment effects.

PROTOCOL:

1. Data sources

To demonstrate the calculations, we used annual data on HCC mortality from 1976 to 2015 for men and women in Taiwan. Statistical package for social sciences (SPSS) version 24.0 for Windows and Microsoft Excel were used to execute the protocols for this study.

1.1 Have the HCC physician classify the patients' clinical symptoms, laboratory tests and medical imaging results to give a diagnosis code according to the International Classification of

Disease (ICD) Code, ICD 150.

1.2 Ensure that the data file (saved as a CSV) contains the year (i.e., period), age, cohort, death number, mid-year population number, and mortality as columns.

1.2.1 Click **File | Import Data | CSV data | Open**. Make sure there is a check in the box next to Read variable names from the first row of the data and click **OK**. Ensure that the data file is imported into SPSS.

1.3 Construct contingency table data crossed by age-period groups through SPSS. In general, we defined row variables as age and column variables as period. If the data featured single period year (or single age year) data, it was necessary to integrate them into a period group (or age group). Then, we cross-tabulated the attitudes to an age group across the survey years.

1.3.1 Click **Analyze | Descriptive Statistics | Crosstabs** and select the age variable in the box next to row(s) and period variable in the box next to column(s). Click **Cells** and make sure that there is a check in the box next to **Observed**. A contingency table of death number (or mid-year population number, or mortality) can be performed in SPSS by the above steps.

1.3.2 Export the contingency table data that were entered in CSV format for analysis through other software. Click **File | Export Data | Ensure Desired Data Format as CSV | Location**. This non editable field displays the safe location for the exported file.

1.3.3 File name: Click **Select** to change the file name.

1.3.4 Export as type: Select a CSV file type from the drop-down menu. Click **variables** to display the available variables and to select the variable tables. By default, all variables from the source data set are retained for the exported file. Researchers can use the tables to specify which source variables to include in the exported file. Click **Export**.

2. Model setting

NOTE: The multistage method was proposed by Keys and Li ⁸ with graphical investigation. A median polish analysis was performed to eliminate the cumulative effects of age and period; finally, these residuals from the median polish phase in the cohort category in the linear regression model were regressed, and cohort effects using data in the contingency table were evaluated.

2.1.1 Graphical representation as the first phase

2.1.1.1 Create a line plot of age groups and period groups. To inspect birth cohorts across age groups or birth cohorts, draw even birth cohorts across ages or cycles in the line charts.

2.1.1.2 Import a CSV file with contingency table mortality data. Click **File | Open | Browse to**

select a CSV file from a folder. Remember to choose **All Files** in the drop-down list next to **File** name box.

2.1.1.3 Click **Open** to open the CSV file. Highlight the rows and columns of the mortality contingency data and click **Insert | Charts | Line Graph**.

2.1.2 Median polish analysis as the second phase

2.1.2.1 Iteratively subtract the median from each row and each column to eliminate the cumulative effect of age and period. After the median polishing phase, keep the residuals for the regression procedure to evaluate the cohort effects.

2.1.2.2 Compute the overall median and residual table. Import a CSV file with contingency table mortality data (refer to 2.1.1).

2.1.2.3 LN was used for each cell of the contingency table mortality data. Click **Formulas | Mathematical & Trigonometry Function** and select **LN**.

2.1.2.4 Number: Enter the location label for each cell. Ensure each cell of the contingency table mortality data took LN. Click **Formulas | More Functions | Statistics** and select **MEDIAN**.

2.1.2.5 Number1: Enter the first cell location label.

2.1.2.6 Number2: Enter the last cell location label. Ensure that the resulting median value is stored in the upper left-hand margin of the contingency table. Ensure a residual table is created by taking the difference between the original value (i.e., the LN mortality data) and the overall median.

2.1.2.7 Compute the row medians (i.e., the medians of each age group) and ensure it computed the row median values for the response age group. Click **Formulas | More Functions | Statistics | Select MEDIAN**.

2.1.2.7.1 Number 1: Enter the first cell location label of the raw sample.

2.1.2.7.2 Number 2: Enter the last cell location label of the raw material. Ensure the resulting row median values are stored in the left-hand margin of the contingency table.

2.1.2.8 Create a new residual table after subtracting from the row medians. Ensure a new set of residual values is created from the row medians where each cell takes on the value of the subtraction of the row median from each response variable in that row. Click **=** and ensure each row's overall cell location label has subtracted the label of the left-hand margin median.

2.1.2.9 Compute the column medians (i.e., the medians of each period group) and ensure it computes the column median values for the response period group. Click **Formulas | More**

Functions | Statistics | Select MEDIAN.

2.1.2.9.1 Number 1: Enter the first cell location label of the column. Number 2: Enter the last cell location label of the column. Ensure the resulting column median values are stored in the upper margin of the contingency table.

2.1.2.10 Create a new residual table after subtracting from the column medians. Ensure a new set of residual values is created from the column medians where each cell takes on the value of the subtraction of the column median from each response variable in that column. Click = and ensure each column overall cell location label has subtracted the label of the upper margin median.

2.1.2.11 Repeat Steps 2.1.2.7 to 2.1.2.10 until the row and column medians approximate zero. Click **Formulas | More Functions | Statistics | Select MEDIAN**. Make sure the row and column medians are approximately zero. Save the final residual table in CSV format.

2.1.3 Regression procedure with weight as the third phase

NOTE: We calculated the dependent variable as residuals for each cohort with death number as the weight. Next, we ran linear regressions to calculate the cohort effects.

2.1.3.1 Ensure Kutools for Excel was installed and use its Transpose Table Dimensions tool to quickly convert cross tables to flat lists. Import a CSV file with contingency table residual data (refer to 2.1.1.1).

2.1.3.2 Select the table to be converted to a list. Click the **Kutools | Modify | Transpose Table Dimension**. In the Transpose Table Dimensions dialog box, make sure there is a check in the box next to the Cross table to list, and select the Results range to store the residuals in list format.

2.1.3.3 Column insert in the initial data file (refer to 1.2) with the residual list format data (refer to 2.1.3.1). Ensure it inserted a supporting column in the residual list format data (refer to 2.1.3.1). Click = age & period variables and click **Enter**. Use a supporting column to look up the age and period group labels of the residual list format data to insert a response residual column in the initial data file (refer to 1.2).

2.1.3.4 Click **Formulas | Lookup & Reference | Select VLOOKUP**. Set VLOOKUP (cell location label of age & cell location label of the period, first cell location label of the supporting column: the last cell location label of the residual column, 4, 0). Make sure the select range includes supporting, age, period, and residual columns (i.e., the 4th column as the residual list).

2.1.3.5 Ensure that the residuals are inserted in the initial data file (refer to 1.2) look up for the residual list format data (refer to 2.1.3.1) for the next step. Fit the regression model by unweighted least squares and analyze the residuals.

2.1.3.6 Click **Analyze | Regression | Linear**. Transfer the independent variable, cohort category (i.e., 17 birth cohorts), into the Independent(s), box and the dependent variable, Residuals, into the Dependent: box. Click **OK**. Ensure generate the results of the unweighted cohort effects.

2.1.3.7 Ensure that the residuals are inserted in the initial Excel data file (refer to 1.2) look up for the residual list format data (refer to 2.1.3.1) for the next step. Fit the regression model by weighted least squares and analyze the residuals. Click **Analyze | Regression | Linear**.

2.1.3.8 Transfer the independent variable and cohort category (i.e., 17 birth cohorts) into the independent(s) box and the dependent variable and residuals into the dependent: box. Transfer the death number into the WLS weight box. Click **OK**. Ensure it generates the results of the weighted average of the cohort effect.

REPRESENTATIVE RESULTS:

The mortality data were demonstrated for 10 five-year age groups (40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, and 85+) and 8 five-year time periods (1976-1980, 1981-1985, 1986-1990, 1991-1995, 1996-2000, 2001-2005, 2006-2010 and 2011-2015). The number of cohort groups was selected by subtracting one from the total number of age-period groups: 10 (five-year age groups) + 8 (five-year time periods) - 1 = 17 birth cohorts, with the birth cohort groups denoted by mid-cohort years as 1891, 1896, 1901, 1906, 1911, 1916, 1921, 1926, 1931, 1936, 1941, 1946, 1951, 1956, 1961, 1966, and 1971. We provide a format of the age group contingency of men and women with HCC (in **Supplementary Table 1**). **Figure 1** and **Figure 2** show the HCC mortality rates within the age and period groups. These fluctuations were more substantial among men than among women. The ratio based on the age distribution shows that at the lower end of the 40-44 age group, the HCC mortality rates are increasing (**Figure 1**). In contrast, **Figure 2** clearly shows that the HCC mortality rates gradually increased in the ≥60 age group. However, the HCC mortality rates based on age have changed substantially over time, which means that a remarkable cohort effect hidden in the normal age-period cross-classified Vital Statistics table will not become apparent until some point in the future.

We implemented the median polishing procedure on log-transformed HCC mortality rates. The estimated cohort effects for the APC model of the HCC mortality rates for men and women are shown in Tables 1 and 2, respectively. In addition, the calculation of the two gender-weighted average procedures before obtaining the weighted estimates is reported in **Tables 1 and 2**. The weighted estimates better conform to the data than the previous estimated cohort effects, based on the minimum deviation (comparing equally weighted estimates) of the confidence interval (CI) of the weighted estimates.

For men, the left panel of **Table 1** shows the cohort effects of the birth cohorts. The cohort effect increases from 0.73 (the earliest cohort effect in 1891) to 1.20 (the greatest cohort effect in 1936). For women, the cohort effect increases from 0.68 (the earliest cohort effect in 1891) to 1.35 (the greatest cohort effect in 1936). It is important to emphasize that compared to the 1891 cohort, the cohort effect for men and women increased by approximately 64% and 98%, respectively. The increase was equally distributed in the right panel of **Table 2**. Here, the cohort

effect increased from 0.71 (the earliest cohort effect in 1891) to 1.11 (the greatest cohort effect in 1936). For women, a similar effect of increased distribution is demonstrated in the right column of Table 2. The cohort effect increased from 0.64 (the earliest cohort effect in 1891) to 1.11 (the greatest cohort effect in 1926). Therefore, compared to the earliest cohort of men and women, we observed an increase in mortality rates of approximately 57% and 73%, respectively.

Among the birth cohorts, men born in approximately 1936 showed the highest risk for HCC mortality (**Supplementary Table 1**). Therefore, for the weighted estimates, the impact of the birth cohort in 1936 compared to that of the reference birth cohort in 1921 was 1.11 (95% CI: 1.08-1.14). In contrast, the earlier cohort of 1891 showed a sharply increasing trend. In addition, the effects were reversed after the 1936 cohort. In Table 1, compared with the reference birth cohort of 1916, the weighted effect was 1.11 (95% CI: 1.07-1.16). In addition, with men and women, we modeled equally weighted and weighted cohort effects, respectively, with 95% confidence intervals. Both figures show that the equally weighted cohort effects are broader than almost all of the widths of the 95% CIs.

FIGURE AND TABLE LEGENDS:

Figure 1. HCC mortality rates per 100,000 by age and period, men, Taiwan, 1976-2015.

Figure 2. HCC mortality rates per 100,000 by age and period, women, Taiwan, 1976-2015.

Figure 3. Age-adjusted mortality rate of death from hepatocellular carcinoma for men and women in Taiwan.

Table 1. Estimated rate ratios and 95% conference intervals for the effect of birth cohort on hepatocellular carcinoma mortality of men in Taiwan, 1891-1971.

Table 2. Estimated rate ratios and 95% conference intervals for the effect of birth cohort on hepatocellular carcinoma mortality of women in Taiwan, 1891-1971.

DISCUSSION:

Due to the time trend of HCC mortality, conventional models underestimate some important features hidden in the data (such as cohort effects), and conventional analyses that use simple linear extrapolation of the observed logarithmic age correction rate show significantly reduced accuracy in their predictions. It is clear that this trend has continued for 35 years and will trend upwards in the next few years if we directly observe the long-term trend of HCC mortality in Taiwan from 1976 to 2015 (Figure 3). Indeed, the recent trend of HCC mortality in Taiwan is declining and is driven by the cohort effect (determined by APC analysis), which, as mentioned earlier, declined after the 1936 cohort. This study shows that the application of the APC model provides advanced and more accurate warnings about trend changes.

From a clinical perspective, there are approximately two billion people infected by hepatitis B virus (HBV) ¹², and approximately 350 million people suffer as a result. Consequently, this is a

significant health problem with high morbidity worldwide. HBV infection causes a wide range of clinical problems, including ineffective carrier status to fulminate hepatitis, cirrhosis, or hepatocellular carcinoma. The most effective prevention method is to inoculate individuals with hepatitis B vaccine. Taiwan implemented the first global hepatitis B mass vaccination plan in 1984¹³. In this program, pregnant women were screened for hepatitis B surface antigen (HBsAg) and hepatitis B envelope antigen (HBeAg)¹⁴. For the first two years of this program, the immunization program only covered babies of mothers with HBsAg. However, beginning in the third year of the vaccination program, all babies were covered. The coverage rate of the hepatitis B vaccine has reached 99% in recent years¹⁵. Nearly 90% to 95% of people will experience lifelong immunity after receiving the three doses of the vaccine. We emphasize that the decline in pediatric HCC in Taiwan is largely attributed to this global vaccination program.

The APC modeling depicted in this article provides advanced warning about these (increased) trend changes (which will decrease in the near future). When comparing the trend of cohort effects (**Tables 1 and 2**) and age-adjusted mortality (**Figure 3**), the direct age-adjusted mortality rate (or age standardization mortality rate) was the same as the weighted average. It weights the age-mortality rates by the proportion of the age group of interest based on the 2000 World Standard Population¹⁶ in this study. As the validation of cohort effects dominated the latest pattern of HCC mortality, we calculated the age-adjusted HCC mortality for the up-to-date data (until 2011-2015). We interpreted this to mean that the weighted mean estimate of the cohorts provides reliable information while research is prepared to forecast future HCC mortality. Details on forecasting HCC mortality are available in our previous study⁵.

A general hypothesis is that each value within the data provides equal information to evaluate the parameters in a model. This approach has been used in most modeling methods (such as linear or nonlinear regression models) and means that the standard deviation of the error term is the constant underlying predictor variable. However, according to our literature review, this hypothesis is not suitable for modeling to empirically estimate parameters. The unknown parameters are estimated when we use weighted regression, which generates a smaller weight with less accuracy for data points and a large weight with more accuracy for data points. The weighting process diminished the standard deviation of the estimator. Nonetheless, the shortcomings of the weighted regression method are almost unknown in empirical practice. Since the exact weights are not known, the estimated weights were used to estimate the parameters. In addition, previous experience has demonstrated that the weighting based on estimation does not significantly change or usually affects regression analysis or its interpretation¹⁷. Hypothetically, the APC model can be fitted to any disease in which the incidence is affected by age, period, and cohort. In addition, weighted mean estimates were made available for prediction¹⁸⁻²⁰. If the CI is comparatively narrow, the uncertainty is small. In view of the fact that the CI depicts the uncertainty inherent in this type of evaluation and the values within, we generally conclude that using CI has a substantial impact.

Transcatheter arterial chemoembolization (TACE) is one of the most efficient methods to clinically control HCC. However, it is difficult to choose this method as the prime or auxiliary therapy, as it does not require open surgery. The liver typically provides 75% of the blood and

nutrients through the hepatic portal vein, while the hepatic artery provides 25% of the blood and nutrients. In contrast to the hepatic artery blood extracted from most HCCs, this fluid increases rapidly and rarely comes from the hepatic portal vein. Moreover, this effect is well suited to TACE because primary liver cancer rarely metastasizes to other parts of the body. Even though hepatocellular malignant tumors are unlikely to metastasize, they are difficult to eradicate. In clinical practice, follow-up for HCC patients is conducted every two to three months. Once abnormal elevation of alpha-fetoprotein (AFP) or an abnormal ultrasound check is detected, computed tomography and magnetic resonance imaging are performed. If a new tumor is discovered, then TACE will be considered. New biomarkers have also been developed to detect the recurrence of HBV-related HCC, such as the HBV DNA quantitative-time index (HDQTI) ²¹. The product of the follow-up results and the logarithm of the detected to normal HBV DNA load ratio is the summation of the HDQTIs. The HDQTI is used as an independent prognostic indicator of HBV-related HCC recurrence ²¹.

Our study has several limitations. First, we merely hypothesized regarding the etiologies of the observed changes. Using the APC model, HCC mortality according to age, period, and cohort effects was reconsidered. Nevertheless, in this study, we used the median polishing setting as an assumption. Second, APC analysis has been widely used in the field of epidemiology in developing or recently developed countries for long-term cohort studies. Third, we did not have information from the accumulated format data set to adjust for confounding factors in the APC model, such as comorbidities or lifestyle. Isolated data are needed for future research to resolve this limitation. Fourth, to modify the regression procedure in the multistage method, we used the number of deaths due to HCC as the weight. Since the exact weight is not known, the use of various weights causes slight inflation within the estimated cohort effects. Ultimately, there are various APC estimation methods to solve the unrecognizable problem (for example, Holford uses linear and curvature trends to solve the unrecognizable problem ²²). At the same time, the median polish provides complex assumptions in the form of conceptual conversion among APC models to evaluate the cohort effect with the fewest assumptions and easily applies a common format for contingency tables.

Overall, the weighted mean effect with a comparatively narrow CI of each cohort was then allowed by the weighted estimation to modify the regression model. Briefly, for multistage methods, it is advisable to use weighted estimates of regression models to evaluate the cohort effects in age-period contingency table data.

ACKNOWLEDGMENTS:

This work was supported by Taipei Tzu Chi Hospital TCRD-TPE-109-39 (1/2) and TCRD-TPE-109-39 (2/2).

DISCLOSURES:

The authors have nothing to disclose.

REFERENCES:

1. Kuntz, E., Kuntz, H.D. *Hepatology: Principles and Practice*. New York, Springer. **774** (2006).

2. McGlynn, K.A. et al. International trends and patterns of primary liver cancer. *International Journal of Cancer*. **94** (2), 290-296 (2001).
3. Bosch, F.X., Ribes, J., Diaz, M., Cleries, R. Primary liver cancer: worldwide incidence and trends. *Gastroenterology*. **127** (suppl), 5S-16S (2004).
4. Tzeng, I.S., Ng, C.Y., Chen, J.Y., Chen, L.S., Wu, C.C. Using weighted regression model for estimating cohort effect in age-period contingency table data. *Oncotarget*. **9** (28), 19826-19835 (2018).
5. Tzeng, I.S., Lee, W.C. Forecasting hepatocellular carcinoma mortality in Taiwan using an age-period-cohort model. *Asia-Pacific Journal of Public Health*. **27**, NP65-73 (2015).
6. Tzeng, I.S. et al. Predicting emergency departments visit rates from septicemia in Taiwan using an age-period-cohort model, 1998 to 2012. *Medicine*. **95**, e5598 (2016).
7. Chen, S.H. et al. Age, Period and Cohort Analysis of Rates of Emergency Department Visits Due to Pneumonia in Taiwan, 1998-2012. *Risk Management and Healthcare Policy*. **13**, 1459-1466 (2020).
8. Keyes, K.M., Li, G. A multiphase method for estimating cohort effects in age-period contingency table data. *Annals of Epidemiology*. **20**, 779-85 (2010).
9. Tukey, J. Exploratory data analysis Reading: MS: Addison-Wesley Publishing Company (1977).
10. Selvin, S. Statistical analysis of epidemiologic data: New York: Oxford University Press (1996).
11. Légaré, G., Hamel, D. An age-period-cohort approach to analyzing trends in suicide in Quebec between 1950 and 2009. *Canadian Journal of Public Health*. **104**, e118-23 (2013).
12. Lavanchy, D. Hepatitis B virus epidemiology, disease burden, treatment, and current and emerging prevention and control measures. *Journal of Viral Hepatitis*. **11**, 97-107 (2004).
13. Chang, M.H. et al. Universal hepatitis B vaccination in Taiwan and the incidence of hepatocellular carcinoma in children. Taiwan Childhood Hepatoma Study Group. *New England Journal of Medicine*. **336**, 1855-9 (1997).
14. Lu, F.T., Ni, Y.H. Elimination of mother-to-infant transmission of hepatitis B virus: 35 years of experience. *Pediatric Gastroenterology, Hepatology & Nutrition*. **23** (4), 311-318 (2020).
15. Chien, Y.C., Jan, C.F., Kuo, H.S., Chen, C.J. Nationwide hepatitis B vaccination program in Taiwan: effectiveness in the 20 years after it was launched. *Epidemiologic Reviews*. **28**, 126-35 (2006).
16. Ahmad, O.B. et al. Age standardization of rates: a new WHO standard. Geneva: World Health Organization, 2005. GPE Discussion Paper Series, 31 (2005).
17. da Silva, C.P., Emídio, E.S., de Marchi, M.R. Method validation using weighted linear regression models for quantification of UV filters in water samples. *Talanta*. **131**, 221-7 (2015).
18. Dawes, R.M. The robust beauty of improper linear models in decision making. *American Psychologist*. **34**, 571-82 (1979).
19. Dawes, R.M., Corrigan, B. Linear models in decision making. *Psychological Bulletin*. **81**, 95-106 (1974).
20. Einhorn, H.J., Hogarth, R.M. Unit weighting schemes for decision making. *Organizational Behavior and Human Performance*. **13**, 171-92 (1975).
21. Wang, W. et al. Association of hepatitis B virus DNA level and follow-up interval with hepatocellular carcinoma recurrence. *JAMA Network Open*. **3** (4), e203707 (2020).
22. Holford, T.R. The estimation of age, period and cohort effects for vital rates. *Biometrics*. **39**,

441 311–324 (1983).

[Click here to access/download;Figure;fig1_revised.tiff](#) 

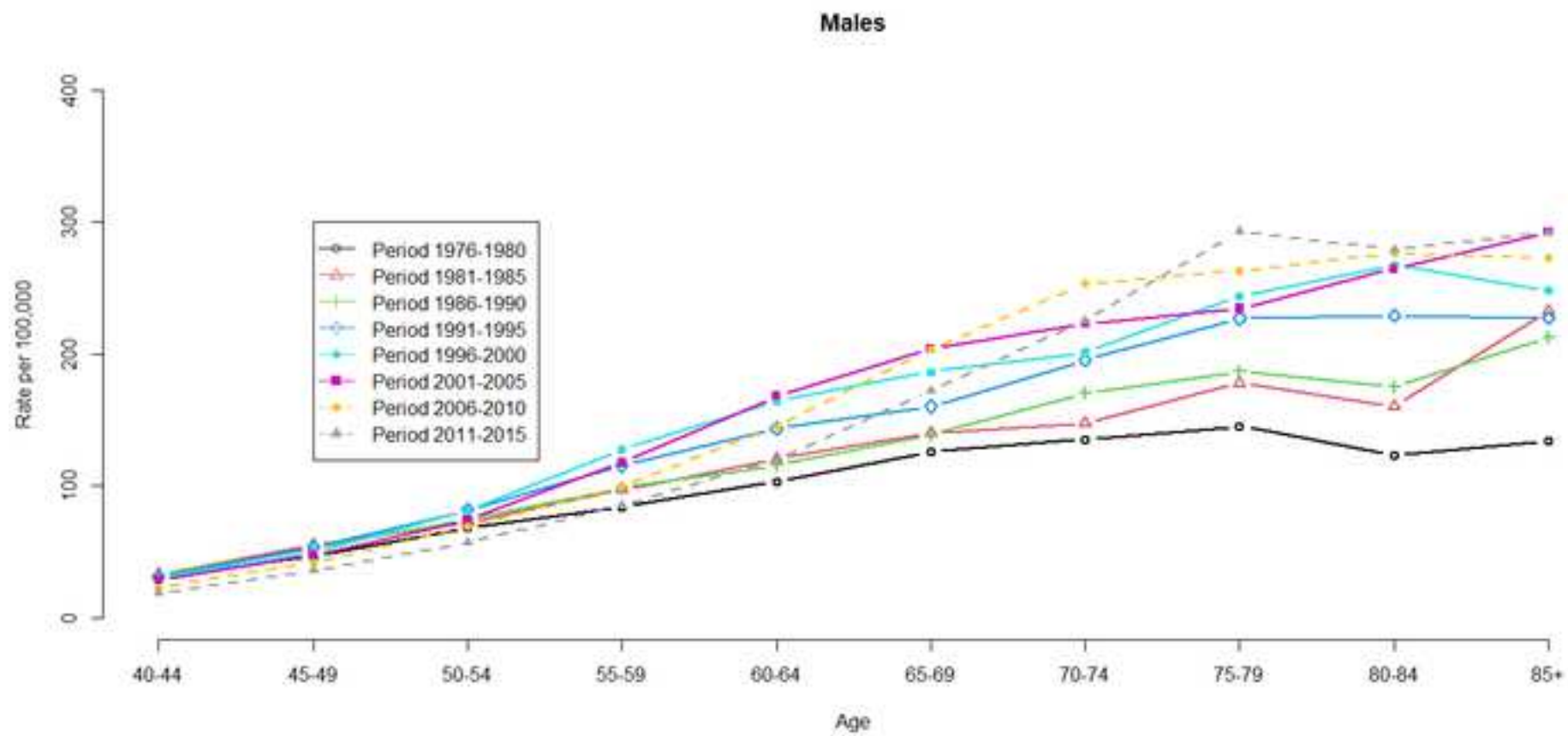


Figure 2

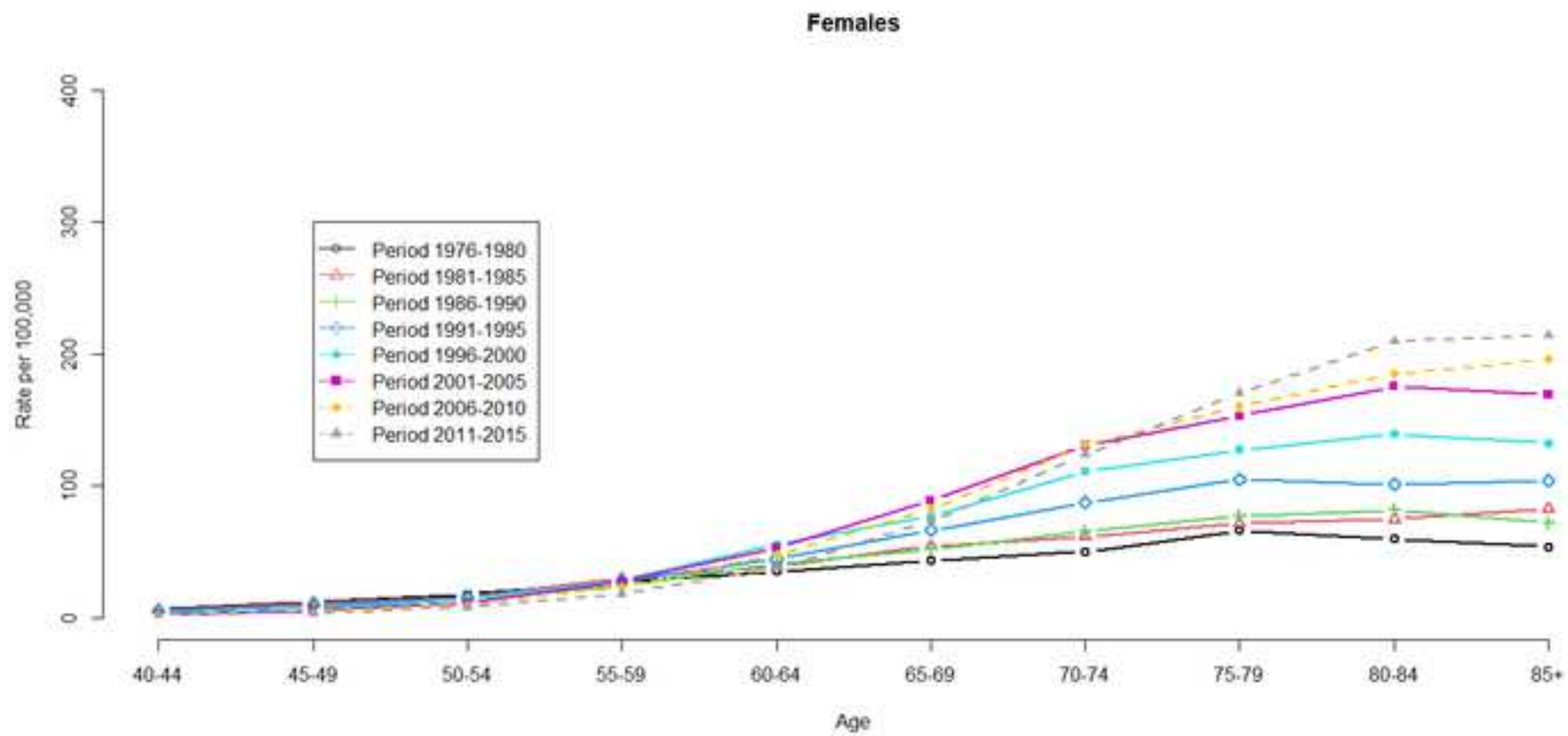


Figure 3

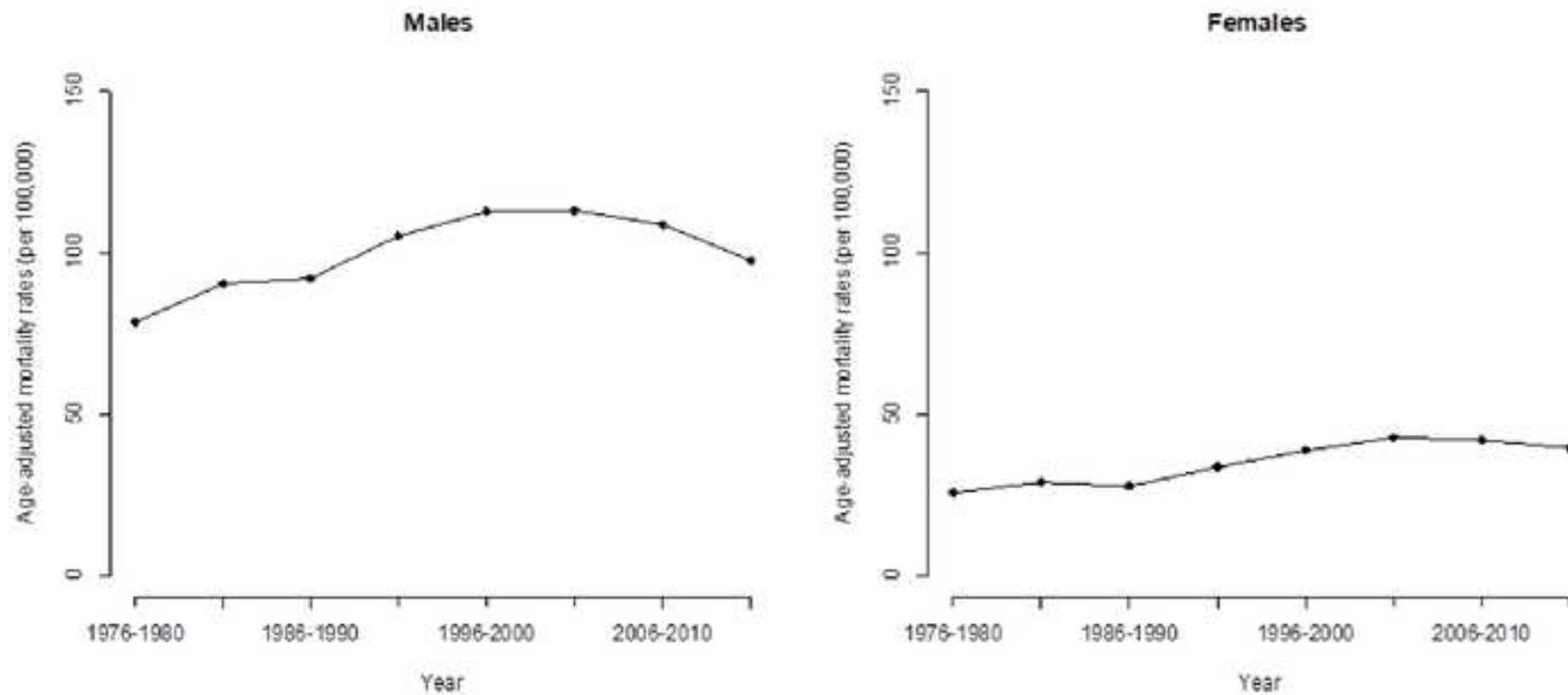


TABLE 1. Estimated rate ratios and 95% confidence intervals for the effect of birth cohort on hepatocellular carcinoma mortality of men in Taiwan, 1891-1971.

Cohort (1891~1971)	Unweighted		Weighted	
	Effects	95% CI for Effects	Effects	95% CI for Effects
1891	0.73	0.59-0.90	0.71	0.57- 0.88
1896	0.88	0.79-0.99	0.87	0.78- 0.97
1901	0.89	0.83-0.96	0.81	0.71- 0.92
1906	0.91	0.86- 0.97	0.85	0.78- 0.94
1911	0.95	0.90-1.00	0.89	0.83- 0.96
1916	1.01	0.97-1.06	0.99	0.95- 1.03
1921	1	REF	1	REF
1926	1.04	1.00-1.08	1.03	1.01- 1.06
1931	1.1	1.06-1.14	1.08	1.06- 1.11
1936	1.2	1.15- 1.24	1.11	1.08- 1.14
1941	1.14	1.09- 1.19	1.1	1.07- 1.13
1946	1.04	1.00-1.09	1.06	1.04- 1.09
1951	0.91	0.87-0.96	1	0.98- 1.03
1956	0.87	0.82-0.92	0.96	0.93- 0.98
1961	0.82	0.76-0.88	0.88	0.85- 0.92
1966	0.76	0.68- 0.85	0.79	0.74- 0.83
1971	0.71	0.57-0.87	0.83	0.80- 0.87

Note: REF = reference; CI = confidence interval.

TABLE 2. Estimated rate ratios and 95% confidence intervals for the effect of birth cohort on hepatocellular carcinoma mortality of women in Taiwan, 1891-1971.

Cohort (1891~1971)	Unweighted		Weighted	
	Effects	95% CI for Effects	Effects	95% CI for Effects
1891	0.68	0.42- 1.10	0.64	0.38-1.09
1896	0.81	0.63-1.04	0.75	0.56- 1.00
1901	0.8	0.67- 0.95	0.7	0.52- 0.94
1906	0.83	0.72- 0.95	0.76	0.65- 0.88
1911	0.88	0.78- 0.99	0.85	0.78- 0.93
1916	1	REF	1	REF
1921	1.12	1.01-1.24	1.08	1.03- 1.13
1926	1.29	1.17-1.42	1.11	1.07- 1.12
1931	1.3	1.18-1.43	1.1	1.05- 1.15
1936	1.35	1.22-1.49	1.1	1.04- 1.14
1941	1.19	1.07-1.32	1.09	1.03-1.13
1946	1.05	0.94-1.17	1.06	1.02-1.11
1951	0.83	0.73-0.94	1	0.96-1.05
1956	0.67	0.58-0.77	0.93	0.89-0.98
1961	0.58	0.49-0.70	0.79	0.74-0.84
1966	0.59	0.46-0.75	0.58	0.49-0.69
1971	0.63	0.40-1.02	0.64	0.58-0.72

Note: REF = reference; CI = confidence interval.



Click here to access/download

Table of Materials
JoVE_Table_of_Materials.xlsx

June 17, 2021

Original Manuscript ID: JoVE62253R2

Original Article Title: Forecasting Hepatocellular Carcinoma Mortality Using a Weighted Regression Model to Estimate Cohort Effects in Taiwan

To: JoVE Editor

Re: Response to Editor

Dear Editor,

My colleagues and I are pleased to learn that you are interested in reviewing a revised version of our manuscript. Thank you for allowing a resubmission of our manuscript, with an opportunity to address the editor and reviewers' comments. We found the editor's comments are reasonable and insightful. We have addressed concerns of the editor and reviewers as detailed in the "response to editor" section attached at the end of the letter. We have also made careful revisions according to the editor and editor's suggestions.

We are uploading (a) our point-by-point response to the comments (below) (response to editor and reviewers) and (b) an updated manuscript with red color indicating changes. We hope the revised manuscript satisfactorily meets the expectations for publication by the Journal. However, not fully answer situation may occur due to comprehension gap between us and editor or reviewers. Limited experienced in protocol draft, we apologized for any inconvenience caused by us. We promised to revise this protocol until suitable for film. We also attached AJE English Editing certificate for reference. We would greatly appreciate any further changes for manuscript presentation if you require again. We look forward to your positive feedback.

Best regards,

Dr. I-Shiang Tzeng

Taipei Tzu-Chi Hospital, Taiwan

Editor, Concern #1. Please revise for clarity.

Editor, Highlight: "From the perspective of the etiologies of alcohol and viral, it is closely related to liver cirrhosis."

Author response: We have revised accordingly.

Revised action: HCC is associated with liver cirrhosis with alcohol and viral etiologies

Editor, Concern #2. Please revise for clarity.

Editor, Highlight: "Since separating the residuals from the median polishing stage, to return these residuals to evaluate their impact on the cohort category is the final stage (cohort is an indicator variable), and the proportion of HCC deaths in each cohort is the same as its influence."

Author response: We have revised accordingly.

Revised texts: The HCC deaths were determined for each cohort with its weighted influence.

Editor, Concern #3. Please provide the terminal commands or the button clicks in the software to perform the actions in the protocol. We need these details in order to film.

Editor, Highlight: "The software SPSS version 24.0 used to execute protocols for this study."

Author response: We thanks for editor's suggestions and comments. We have added accordingly.

Editor, Concern #4. How are the identifications and classifications done?

Editor, Highlight: "we identify and group classification of cause of death into specific disease categories using the International Classification of Disease (ICD) Code. HCC mortality was classified according to ICD 150."

Author response: We thanks for editor's suggestions and comments. We examined the physician identified patients' clinical symptom, lab test and medical image to give diagnosis identifications according to International Classification of Disease (ICD) Code.

Revised texts: Have the HCC physician classify the patients' clinical symptoms, laboratory tests and medical imaging results to give a diagnosis code according to the International Classification of Disease (ICD) Code, ICD 150.

Editor, Concern #5. What is done on this website? What data is downloaded specifically?

Editor, Highlight: "Go to <https://dep.mohw.gov.tw/DeptList.html>. We gained this data

from individual health records of the Ministry of Health and Welfare (MOHW), Taiwan.”

Author response: We thanks for editor’s suggestions and comments. We examined we download HCC death number among 0, 1, 2, 3, 4, 5-9, .., 85+ age groups annually (1976-2015) on this website. We also download mid-year population number on this website. For consideration, we focused on protocol tutorial after researcher prepared a data file contained Year (i.e., Period), Age, Cohort, Death number, Mid-year population number, and mortality as columns. (i.e., Death number divided by Mid-year population number as HCC mortality).

Editor, Concern #6. How is this specifically done? You have specified what to do but we need to know how to do it.

Editor, Highlight: “Construct contingency table data crossed by age-period groups. In general, we defined row variable as age and column variable as period. If the data featured as single period year (or single age year) data, it is needed to integrate as period group (or age group). The mortality data were useful for 10 five-year age groups (40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, and 85+), 8 five-year time periods (1976-1980, 1981-1985, 1986-1990, 1991-1995, 1996-2000, 2001-2005, 2006-2010 and 2011-2015).”

Author response: We thanks for editor’s suggestions and comments. We examined that researcher determined risk age ranges and covered period based on previous empirical studies. To our knowledge, HCC mortality highly associated with age so that collected data among 45-49 to 85+ age groups and covered period from 1976 to 2015.

Editor, Concern #7. How is this specifically done? You have specified what to do but we need to know how to do it.

Editor, Highlight: “The number of cohort group is decided by minus one from the total number of age-period groups. 10 (five-year age groups) +8 (five-year time periods) -1 = 17 birth cohorts, those birth cohort groups denoted by mid-cohort years as 1891, 1896, 1901, 1906, 1911, 1916, 1921, 1926, 1931, 1936, 1941, 1946, 1951, 1956, 1961, 1966, and 1971.”

Author response: We thanks for editor’s suggestions and comments. Note that we actually have two timescales: Time since birth (i.e. since entry into the study) and Calendar time (i.e., dead). These can be shown simultaneously in a Lexis diagram. That’s why we had $10+8-1=17$ birth cohorts.

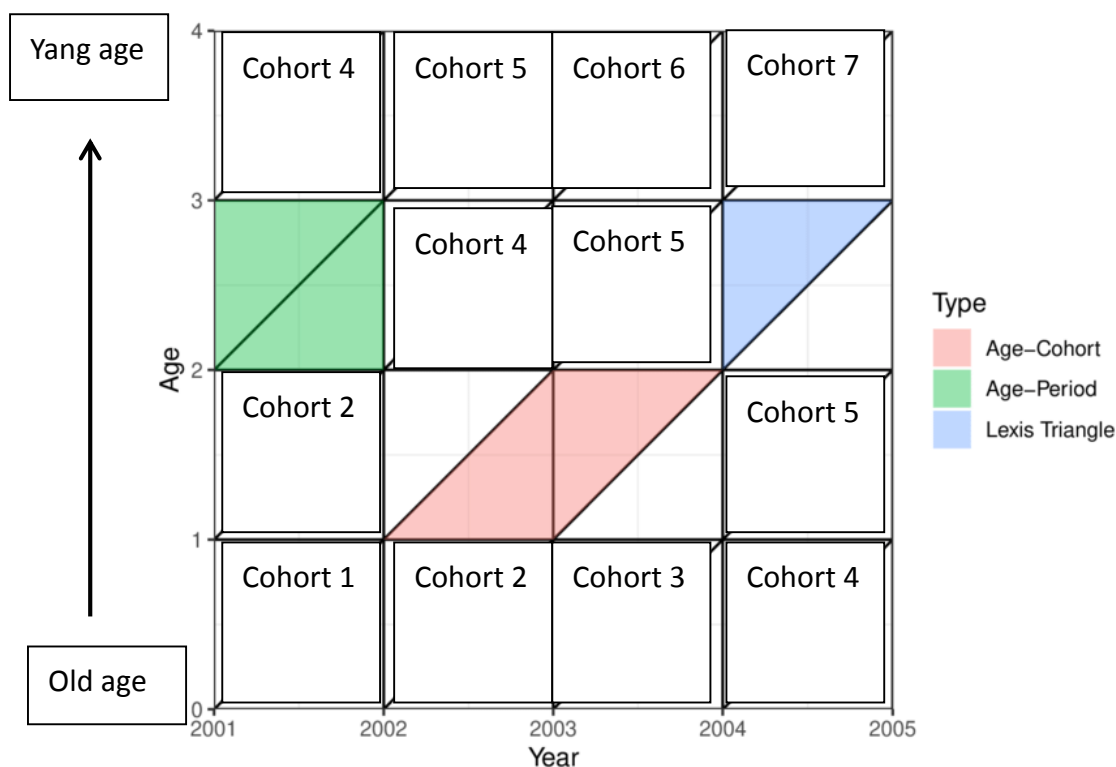


Illustration of lexis diagram

(For example, 4 age groups and 4 period groups, then $4+4-1=7$ cohort groups)

Cite:

https://www.researchgate.net/publication/335976100_Forecasting_Fertility_with_Parametric_Mixture_Models/figures?lo=1&utm_source=google&utm_medium=organic

Editor, Concern #8.1 How is this specifically done? You have specified what to do but we need to know how to do it.

Editor, Highlight: “APC model estimates patterns in cancer mortality rates from population-based Count (numerator) and Population (denominator) data. Often the data come from a Cause of Death Statistics in the form of a table showing the numbers of cancer deaths (counts) and corresponding person-years at risk (population) for particular age groups and calendar time periods (i.e., annual period). In cancer research, the Age-Period-Cohort (APC) framework is a fundamental model to analyze these data. The APC Model includes parameters that describe the mathematical relationships between the mortality rate of cancer and corresponded age, calendar period (year of diagnosis), and birth cohort (year of birth). The cancer mortality rate is expressed as the number of cancers per 100,000 persons at risk, which is calculated from the data using the formula: $(\text{Count}/\text{Population}) \times 100,000$.”

Author response: We thanks for editor’s suggestions and comments. The frame work of APC model setting referred to our previous study [ref#5: Tzeng IS, Lee WC.

Forecasting hepatocellular carcinoma mortality in Taiwan using an age-period-cohort model. Asia Pac J Public Health. 2015; 27:NP65-73.]. The cancer mortality rate is expressed as the number of cancers per 100,000 persons at risk. Let cancer deaths (Count) divided by population number and then multiple 100,000 (i.e., calculated from the data using the formula: $(\text{Count}/\text{Population}) \times 100,000$). For consideration, we focused on protocol tutorial after researcher prepared a data file contained Year, Age, Death number, Mid-year population number, and mortality.

Editor, Concern #8.2 How is this specifically done? You have specified what to do but we need to know how to do it.

Editor, Highlight: "Adopt the multi-stage method to evaluate cohort effect

The multi-stage method was put forward to evaluate the cohort effect 8. According to the multi-stage method, there are three stages to evaluate the cohort effect under the assumption of minimum data in the contingency table."

Author response: We thanks for editor's suggestions and comments. We examined the multi-stage method was proposed by Keys and Li 8 with graphical investigation; next, a median polish analysis to eliminate the cumulative effects of age and period; final, regress these residuals from the median polish phase on the cohort category in the linear regression model and evaluate cohort effects using data in the contingency table. We revised accordingly.

Editor, Concern #8.3 How is this specifically done? You have specified what to do but we need to know how to do it.

Editor, Highlight: "Graphical representation as first phase

Birth cohorts across age groups or birth cohorts, even birth cohorts across ages or cycles were drawn in charts. For instance, we carried out charts of ages in various periods. There were period effects in the contingency table data if the specific age ratios of different age groups varied in parallel among different periods. If the age characteristics rates of different age groups influence each other in different periods, then cohort effects existed."

Author response: We thanks for editor's suggestions and comments. We examined conduct line plot of age groups and period groups through Excel. To inspect birth cohorts across age groups or birth cohorts, even birth cohorts across ages or cycles were drawn in line charts. We revised accordingly.

Editor, Concern #9. What data is entered? How is the excel spreadsheet set up?

Editor, Highlight: "Enter the data in Excel. Also, the data can be imported into Excel from other applications."

Author response: We thanks for editor's suggestions and comments. We examined import a CSV file with a contingency table mortality data (refer to 2.1.1.2) by opening it in Excel. We revised accordingly.

Editor, Concern #10. Specify the data that is used.

Editor, Highlight: "Select the data that will be used to create a graph. Don't forget to include the labels (i.e., overall age and period groups) as well."

Author response: We thanks for editor's suggestions and comments. We examined import a CSV file with a contingency table mortality data (refer to 2.1.1.2) by opening it in Excel. We revised accordingly.

Editor, Concern #11. What data is selected? We need this level of details throughout: Click Insert | Charts | Line Graph

Editor, Highlight: "When the required data have been selected, click Insert => Charts=>Line Graph"

Author response: We thanks for editor's suggestions and comments. We examined import a CSV file with a contingency table mortality data (refer to 2.1.1.2) by opening it in Excel. We revised accordingly.

Editor, Concern: #12.1. How is this specifically done? You have specified what to do but we need to know how to do it.

Editor, Highlight: "Median polish analysis as second phase

Without depending on a specific distribution, it is the benefit of median polish. Median polishing was widely applied to various types of data, such as ratios, log ratios, ratios and counts. Iteratively subtract the median from each row and each column to eliminate the cumulative effect of age and period."

Author response: We thanks for editor's suggestions and comments. We revised accordingly. We examined iteratively subtract the median from each row and each column to eliminate the cumulative effect of age and period. After median polish phase, the residuals kept for the regression procedure to evaluate cohort effects.

Editor, Concern: #12.2. How is this specifically done? You have specified what to do but we need to know how to do it.

Editor, Highlight: "Compute overall median and residual table."

Author response: We thanks for editor's suggestions and comments. We examined import a CSV file with a contingency table mortality data by opening it in Excel (refer to 2.1.1.2). Take LN for each cell of contingency table mortality data. We revised accordingly.

Editor, Concern: #12.3. How is this specifically done? You have specified what to do but we need to know how to do it.

Editor, Highlight: "Compute the row medians (i.e., medians of each age group)."

Author response: We thanks for editor's suggestions and comments. We examine in addition to computing the row median values for the response age group. We revised accordingly.

Editor, Concern: #12.4. How is this specifically done? You have specified what to do but we need to know how to do it.

Editor, Highlight: "Create a new residual table after subtracting from the row medians."

Author response: We thanks for editor's suggestions and comments. We examined a new set of residual values is created from the row medians where each cell takes on the value of the subtraction of the row median from each response variable in that row. We revised accordingly.

Editor, Concern: #12.5. How is this specifically done? You have specified what to do but we need to know how to do it.

Editor, Highlight: "Compute the column medians (i.e., medians of each period group)."

Author response: We thanks for editor's suggestions and comments. We examined in addition to computing the column median values for the response period group. We revised accordingly.

Editor, Concern: #12.6. How is this specifically done? You have specified what to do but we need to know how to do it.

Editor, Highlight: "Create a new residual table after subtracting from the column medians."

Author response: We thanks for editor's suggestions and comments. We examined a new set of residual values is created from the column medians where each cell takes on the value of the subtraction of the column median from each response variable in that column. We revised accordingly.

Editor, Concern: #12.7. How is this specifically done? You have specified what to do but we need to know how to do it.

Editor, Highlight: "Repeated Step 2 to 5 until the row and column medians approximate zero."

Author response: We thanks for editor's suggestions and comments. We revised

accordingly.

Editor, Concern: #12.8. How is this specifically done? You have specified what to do but we need to know how to do it.

Editor, Highlight: "Regression procedure with weight as third phase

In the wake of the median polishing, the residuals include cohort effect and random error. The summary data were the count data in contingency table format, which we regress these residuals on the cohort category in the linear weight regression model. We supposed weight as the number of deaths from HCC in each cohort. Let W_k represent the weight k th cohort category.

$$\varepsilon_k = W_k \times \gamma_k + \varepsilon_{ijk}$$

The regression equation applied each of these weights. A weight equal to the number of HCC deaths in each cohort performed the weighted average of the cohort effect.

We conduct dependent variable as residuals for each cohort and design matrix with death number as weight for each cohort. Next, we run the linear regression to calculate cohort effects."

Author response: We thanks for editor's suggestions and comments. We examined conduct dependent variable as residuals for each cohort with death number as weight. For consideration, we focused on protocol tutorial for researcher prepared a data file contained Year (i.e. Period), Age, Death number, Mid-year population number, mortality, and residuals.

Editor, Concern #13. We need this level of details throughout: Click Insert | Charts | Line Graph

Editor, Highlight: "Fit the regression model by unweights least squares and analyzes the residuals. Click Analyze=> Regression => Linear"

Author response: We thanks for editor's suggestions and comments. We revised accordingly.

Editor, Concern #14. We need this level of details throughout: Click Insert | Charts | Line Graph

Editor, Highlight: "Transfer the independent variable, cohort category (i.e., 17 birth cohorts), into the Independent(s): box and the dependent variable, Residuals, into the Dependent: box. Click on the OK button. This will generate the results of unweight cohort effects."

Author response: We thanks for editor's suggestions and comments. We examined residuals insert in initial Excel data file (refer to 1.2) look up to residual list format

data (refer to 2.1.3.1) for next step. Fit the regression model by unweights least squares and analyzes the residuals. We revised accordingly.

Editor, Concern #15. How is this done?

Editor, Highlight: "Estimate the regression coefficients as weighted average of the cohort effect using setting weights"

Author response: We thanks for editor's suggestions and comments. We revised accordingly.

Editor, Concern #16. We need this level of details throughout: Click Insert | Charts | Line Graph

Editor, Highlight: "Transfer the number of HCC deaths, in to the WLS Weight: box. Click on the OK button. This will generate the results of weighted average of the cohort effect."

Author response: We thanks for editor's suggestions and comments. Ensure residuals insert in initial Excel data file (refer to 1.2) look up to residual list format data (refer to 2.1.3.1) for next step. Fit the regression model by unweights least squares and analyzes the residuals. We revised accordingly.



Editing Certificate

This document certifies that the manuscript

**Forecasting Hepatocellular Carcinoma Mortality Using a Weighted Regression Model
to Estimate Cohort Effects in Taiwan**

prepared by the authors

I-Shiang Tzeng, Chan-Yen Kuo, Chia-Chi Wang

was edited for proper English language, grammar, punctuation, spelling, and overall style
by one or more of the highly qualified native English speaking editors at AJE.

This certificate was issued on **June 14, 2021** and may be verified
on the [AJE website](https://www.aje.com) using the verification code **DE40-D8CD-CF44-414C-D4C5**.



Neither the research content nor the authors' intentions were altered in any way during the editing process. Documents receiving this certification should be English-ready for publication; however, the author has the ability to accept or reject our suggestions and changes. To verify the final AJE edited version, please visit our verification page at [aje.com/certificate](https://www.aje.com/certificate). If you have any questions or concerns about this edited document, please contact AJE at support@aje.com.

AJE provides a range of editing, translation, and manuscript services for researchers and publishers around the world.
For more information about our company, services, and partner discounts, please visit [aje.com](https://www.aje.com).

SUPPLEMENTARY TABLE 1.

Age-period contingency table of HCC mortality rate per 100,000
among males and females, Taiwan, 1976-2015

Males	1976-1980	1981-1985	1986-1990	1991-1995	1996-2000	2001-2005	2006-2010	2006-2015
40-44	31.41	33.1	33.2	31.24	30.4	29.1	23.26	19.3
45-49	46.5	55.13	52.88	53.61	49.79	47.66	42.25	36.25
50-54	68.11	72.47	74.56	81.73	81.65	74.19	69.12	56.96
55-59	84.12	97.32	98.96	115.57	127.55	117.7	100.19	85.59
60-64	103.58	120.45	115.71	143.74	164.14	168.2	145.35	119.83
65-69	126.15	140.12	138.76	160.29	186.58	204.31	203.5	172.15
70-74	135.44	147.79	170.44	195.56	201.4	223	253.46	225.23
75-79	145.41	178.22	186.7	226.86	243.87	234.63	263.24	293.21
80-84	123.63	160.33	175.09	229.23	267.55	264.61	276.39	278.95
85+	133.97	232.56	212.33	227.57	248.07	292.51	272.51	291.86
Females								
40-44	6.75	6.14	4.9	5.11	3.46	2.88	3.03	2.44
45-49	11.99	11.1	9.21	9.76	6.49	5.76	4.9	3.79
50-54	18.2	16.23	14.78	15.53	14.15	11.78	10.2	8.17
55-59	27.57	30.06	24.45	28.43	28.65	28.04	23.68	18.34
60-64	35.23	39.48	41.09	44.75	55.75	52.97	48.01	37.94
65-69	43.49	54.68	52.09	66.36	76.92	89.05	82.8	73.44
70-74	50.4	61.49	65.48	87.47	111.46	130.6	131.32	123.96
75-79	66.47	72.02	77.57	105.09	127.37	153.26	160.98	170.88
80-84	60.12	74.79	81.83	101.26	139.41	175.41	184.9	209.8
85+	54.22	82.91	71.93	103.98	132.42	169.12	196.03	214.09



Editing Certificate

This document certifies that the manuscript

Forecasting Hepatocellular Carcinoma Mortality Using a Weighted Regression Model to Estimate Cohort Effects in Taiwan

prepared by the authors

I-Shiang Tzeng, Chan-Yen Kuo, Chia-Chi Wang

was edited for proper English language, grammar, punctuation, spelling, and overall style by one or more of the highly qualified native English speaking editors at AJE.

This certificate was issued on **June 14, 2021** and may be verified on the [AJE website](#) using the verification code **DE40-D8CD-CF44-414C-D4C5**.



Neither the research content nor the authors' intentions were altered in any way during the editing process. Documents receiving this certification should be English-ready for publication; however, the author has the ability to accept or reject our suggestions and changes. To verify the final AJE edited version, please visit our verification page at [aje.com/certificate](#). If you have any questions or concerns about this edited document, please contact AJE at support@aje.com.

AJE provides a range of editing, translation, and manuscript services for researchers and publishers around the world.

For more information about our company, services, and partner discounts, please visit [aje.com](#).