SAS® GLOBAL FORUM 2020 #SASGF

Paper 5162-2020

Piecewise Latent Growth Curve Models to Test for Discontinuities in Disease Prevalence Trends

Lin Yan, Stephanie Goguen, Lisa M. Lix, University of Manitoba

ABSTRACT

The International Classification of Diseases (ICD) is a standard coding and classification system of diagnoses and health conditions developed by the World Health Organization (WHO). The ICD has changed over time; different versions of the ICD are used in health data repositories of hospital and physician service records that contain diagnosis codes. These changes can result in discontinuities in prevalence estimates for health conditions, such as hypertension, arthritis, and dementia, as the meaning of a condition changes over time. Our research purpose is to demonstrate piecewise latent growth curve models to test for discontinuities in ICD-coded data. We apply these models to health data from one Canadian province (population 1.2 million) for cardiorespiratory health conditions captured in ICD versions 8, 9, and 10. Two transition points are considered as a potential source of discontinuity: the transition from ICD-8 to ICD-9 (Clinical Modification) in 1980 in both hospital and physician records and the transition from ICD-9 to ICD-10 in 2004 in hospital data. We investigated piecewise latent growth curve models with the following characteristics: linear or nonlinear trend within each time segment before and after the transition points, discontinuity with or without an elevation (i.e., change in slope) at the transition point. We also show how SAS® generalized linear mixed model procedures can be used to fit piecewise latent growth curve models and identify the best fitting model. Testing for unexpected changes in chronic disease prevalence estimates over time is important to produce accurate regional and national chronic disease surveillance.

INTRODUCTION

Disease prevalence studies are very important for researchers, guideline developers and policymakers to understand the burden of disease and identify priorities for the delivery of healthcare services, as well as plan disease prevention and health promotion policies. Data on disease prevalence rates and time trends are needed for the development of health economic models and the assessment of interventions.

Administrative hospital data and physician service (i.e., billing) claims data are now commonly used to study disease prevalence and to investigate and monitor disease trends. The World Health Organization's (WHO) *International Classification of Diseases* (ICD) system is a standard coding and classification system of diagnoses that has been used internationally to record diagnoses in administrative health data. Revisions of ICD are produced on a regular basis by the WHO. These revisions may result in the introduction of new codes and changes in the interpretation of existing codes (Janssen & Kunst, 2004). These changes can result in inconsistencies and abrupt changes in disease trend estimates over time. Therefore, assessment of trends in disease prevalence rates is important to assess the potential impact of changes in coding on the accuracy of the estimates that are produced.

The WHO has made major revisions in the content and structure of the ICD. For example, in the Canadian province of Manitoba, diagnoses in hospital abstracts have undergone three changes. First they were coded using the ICD, Adapted, Eighth Revision (ICDA-8) (used from January 1, 1970 to March 31, 1979), then there was a transition to the Ninth Revision, Clinical Modification (ICD-9-CM), which was in use from April 1, 1979 to March 31, 2004. Hospital data are currently coded using the Tenth Revision, Canadian version (ICD-10-CA); this

version of ICD was introduced in 2004. Diagnoses in physician service claims were coded using ICDA-8 from March 1, 1972 to February 28, 1979 and then using ICD-9-CM from March 1, 1979 onward. These changes can result in discontinuities in trends in the number of diagnosed disease cases (e.g., hypertension, arthritis, and dementia), as well as in cause-specific mortality and healthcare use.

Janssen and Kunst (2004) reported that 10.8% of ICD revisions to cause-specific mortality codes were associated with significant discontinuities. Specifically, they found that coding changes between and within revisions of ICD can substantially bias long-term trends in cause-specific mortality. Therefore, they recommended that researchers should evaluate and adjust discontinuities caused by ICD coding changes when describing and analyzing trends.

Slavlova et al. (2018) used a segmented regression analysis to estimate the effect of coding changes from ICD-9 to ICD-10 on injury hospitalization trends. They found that there was a significant immediate change in the percentage of injury hospitalizations coded for unintentional (3.34%) and undetermined intent (-3.39%), and significant change in slope after the transition (without immediate level change) for assault, firearm, cut/pierce, and motor vehicle traffic rates. However, their analysis was based on 6 years' of time series data only, therefore, they could not examine possible variations in level and trend among different groups (such as different age groups).

The purpose of this paper is to demonstrate how to use piecewise latent growth curve models to assess the impact of ICD transitions on disease trends and its variation among different groups. We illustrate these models with health administrative data from one Canadian province (population 1.2 million) for two chronic diseases: hypertension and chronic obstructive pulmonary disease (COPD). The GLIMMIX procedure in SAS was used to fit these piecewise growth curve models and identify the best fit model. Our emphasis is placed on the comparisons among these models in the prediction of disease prevalence rate and trends over time.

METHOD

DATA SOURCES AND STUDY POPULATION

This study is a part of a population-based family health history study conducted in the Canadian province of Manitoba using linked administrative databases, including hospital discharge abstracts, physician service claims, and population registry records. All these databases were from the Manitoba Population Research Data Repository that are held within the Manitoba Centre for Health Policy at the University of Manitoba. These databases can be linked using a unique and scrambled personal health number. Family members in Manitoba can be linked through a 6-digit unique family registration number in the provincial health registry.

The study population was composed of Manitoba residents eligible to receive health insurance benefits who were aged 40 years and older between April 1, 1997 and December 31, 2015 and who could be linked to at least one parent. We excluded individuals who were: (1) offspring without at least 3 years of healthcare coverage prior to the study index date, and (2) offspring with less than 3 months of follow-up after the index date unless due to death or out-of-province migration. The index date for the offspring was April 1, 1997 or the April 1 of the year when the offspring turned 40 years old, whichever came later. A fiscal year extends from April 1 to March 31. Both the offspring and their linked parents were included in the study population.

STUDY COHORTS AND HEALTH CONDITION CASE DEFINITIONS

Separate population-based study cohorts were developed for each year of the observation period. A person eligible to enter a cohort met the following criteria: (1) aged ≥40 years at the beginning of that year; and (2) at least one day of coverage within three years prior to the end date of that year.

We investigate a total of 25 chronic health conditions to test for discontinuities in disease prevalence trends. Two chronic health conditions are presented here for demonstration purposes: hypertension and chronic obstructive pulmonary disease (COPD).

We used ICD diagnoses in hospital abstracts and physician claims to define the health conditions.

Table 1 describes the detailed case definitions and ICD-8, ICD-9-CM, and ICD-10-CA codes used to ascertain cases of hypertension and COPD. We used any diagnostic field in Manitoba hospital discharge records (H) or outpatient physician billing records (P) for identifying the relevant diagnosis codes.

We estimated the prevalence of each heath condition based on data from 1972 to 2015.

Chronic Disease	Case Definition	ICD-8	ICD-9-CM	ICD-10-CA
Hypertension	1+H or 2+P in 3 years	401	401	I10-I13, I15
COPD	1+H or 2+P in 3 years	491, 492	491, 492, 496	J40-44

Table 1. Case Definitions and ICD-8, ICD-9-CM, and ICD-10-CA Codes for Hypertension and COPD

STUDY VARIABLES

Aggregate data were used to examine trends in disease prevalence and impact of ICD version changes on the trend. Table 2 describes the variables used in data aggregation and analyses. Mean age was calculated for each age group and for each year (1972 to 2015) for individual-level age data. This mean age variable was treated as a continuous variable in the statistical analyses. Centered age was used in the analyses for clear interpretation of regression intercepts. Age was centered at approximately 50, which is the mean age of the cohort in 1972.

Variables	Labels	Coding Rules
SEX	Biological Sex	F: Female, M: Male
AGEGRP	Age Group	40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70+
AGE	Mean age for each age group in each year	Centered at 50 years old
Y_1	Year	Takes values of 0-43 for years from 1972 to 2015
Y_2	Year After 1980	Takes value of zero for years from 1972 to 1979, 0-35 from 1980 to 2015
Y_3	Year After 2005	Takes value of zero for years from 1972 to 2005, 0-10 from 2005 to 2015
NUMHYPERTEN	The number of prevalent cases of hypertension	Counts
NUMCOPD	The number of prevalent cases of COPD	Counts
POPULATION	The size of population at risk in each year and each age and sex group	Counts

Table 2. Variables Used in the Data Analysis

MODELS

Aggregated data were stratified by age and sex (yearly data from 1972 to 2015), giving 14 age and sex groups. This type of data set often is referred to as pooled data or panel data, describing each of a number of cross-sectional entities (groups), across a sequence of time periods.

The dependent variable was the observed count of cases (e.g., hypertension, COPD) at each year for each age and sex group. Though Poisson regression is often used for modeling count data, our preliminary analysis indicated that the conditional variance exceeds the conditional mean (Dean, 1992). Thus, negative binomial regression was more appropriate for modeling our outcome (Hilbe, 2011).

Preliminary analyses also indicated that the log of prevalence rates is not a linear function of time for most chronic health conditions. We specified the log of prevalence rates as a quadratic function of time. To examine the impact of ICD code changes on the prevalence rates, we compared three different discontinuities each ICD code change: an immediate shift in elevation but no shift in slope, an immediate shift in slope but no shift in elevation, and immediate shifts in both elevation and slope. We found that for most chronic health conditions, the model with ICD impact on an immediate shift in slope makes most accurate predictions. Therefore, our overall growth trajectories included three distinct segments of time: one segment that describe the trend of log of prevalence rates from 1972 to 1979, one segment that characterizes the trend from 1980 to 2005, and third segment that characterizes the trend from 2006 to 2015. Thus, this piecewise growth curve model has three segments.

We present three different piecewise latent growth curve models (Bollen & Curan, 2006; Crudeck & Klebe, 2002; Kohli & Harring, 2013; Preacher, Wichman, MacCallum, & Briggs, 2008). For Model A, we ignore the longitudinal data or panel data structure and fit the conventional negative binomial regression. In Model B, we take into account the panel data structure and added a random intercept to Model A. In Model C, we specify both intercept and linear slope of time as random effects (Leite & Stapleton, 2011; Singer & Willett, 2003).

Model A

The specification of Model A is as below

$$\log\frac{\mu_{ij}}{POP_{ij}} = \begin{cases} \beta_0 + \beta_1 \mathrm{Y}_1 + \beta_2 \mathrm{Y}_1^2 + \gamma_1 \mathrm{SEX} + \gamma_2 \mathrm{C\_Age} & \mathrm{Year} < 1980 \\ \beta_0 + \beta_1 \mathrm{Y}_1 + \beta_2 \mathrm{Y}_1^2 + \beta_3 Y_2 + \gamma_1 \mathrm{SEX} + \gamma_2 \mathrm{C\_Age} & 1980 \le \mathrm{Year} < 2005 \\ \beta_0 + \beta_1 \mathrm{Y}_1 + \beta_2 \mathrm{Y}_1^2 + \beta_3 Y_2 + \beta_4 Y_3 + \gamma_1 \mathrm{SEX} + \gamma_2 \mathrm{C\_Age} & \mathrm{Year} \ge 2005 \end{cases}$$

with

$$\Pr(Y = y_{ij} \mid \mu_{ij}, \alpha) = \frac{\Gamma(y_{ij} + \alpha^{-1})}{\Gamma(\alpha^{-1})\Gamma(y_{ij} + 1)} \left(\frac{1}{1 + \alpha\mu_{ij}}\right)^{\alpha^{-1}} \left(\frac{\alpha\mu_{ij}}{1 + \alpha\mu_{ij}}\right)^{y_{ij}}$$

$$\alpha = \frac{1}{\nu}$$

where y_{ij} is case count for group *i* at year *j*, μ_{ij} is the mean of y_{ij}, and v is the scale parameter.

From 1972, μ_{ij} is determined by the population for group *i* at year *j* and a quadratic function of year since 1972 as well as the covariates of sex and centered mean age for each age group in each year. From 1980, in addition to above prediction, μ_{ij} is determined by adding a shift

in the linear slope of time (Y2). From 2005 onwards, μ_{ij} is determined by adding another shift in linear slope of time (Y3).

Model B

For Model B, we assume the intercept in Model A varies across groups. The specification of Model B is

$$\log\frac{\mu_{ij}}{POP_{ij}}=\begin{cases}\beta_0+u_{0i}+\beta_1\mathrm{Y}_1+\beta_2\mathrm{Y}_1^2+\gamma_1\mathrm{SEX}+\gamma_2\mathrm{C\_Age} & \mathrm{Year}<1980\\ \beta_0+u_{0i}+\beta_1\mathrm{Y}_1+\beta_2\mathrm{Y}_1^2+\beta_3Y_2+\gamma_1\mathrm{SEX}+\gamma_2\mathrm{C\_Age} & 1980\le\mathrm{Year}<2005\\ \beta_0+u_{0i}+\beta_1\mathrm{Y}_1+\beta_2\mathrm{Y}_1^2+\beta_3Y_2+\beta_4Y_3+\gamma_1\mathrm{SEX}+\gamma_2\mathrm{C\_Age} & \mathrm{Year}\ge2005\end{cases}$$

with

$$\Pr(Y=y_{ij}\mid\mu_{ij},\alpha)=\frac{\Gamma(y_{ij}+\alpha^{-1})}{\Gamma(\alpha^{-1})\Gamma(y_{ij}+1)}\left(\frac{1}{1+\alpha\mu_{ij}}\right)^{\alpha^{-1}}\left(\frac{\alpha\mu_{ij}}{1+\alpha\mu_{ij}}\right)^{y_{ij}}$$

$$u_{0i}\sim N(0,\sigma_0^2)$$

$$\alpha=\frac{1}{\nu}$$

Model C

For Model C, we assume both intercept and linear slope of time varies across the groups. This can be done by specifying them as random effects. The specification of Model C is

$$\log\frac{\mu_{ij}}{POP_{ij}}=\begin{cases}\beta_0+u_{0i}+(\beta_1+u_{1i})\mathrm{Y}_1+\beta_2\mathrm{Y}_1^2+\gamma_1\mathrm{SEX}+\gamma_2\mathrm{C\_Age} & \mathrm{Year}<1980\\ \beta_0+u_{0i}+(\beta_1+u_{1i})\mathrm{Y}_1+\beta_2\mathrm{Y}_1^2+\beta_3Y_2+\gamma_1\mathrm{SEX}+\gamma_2\mathrm{C\_Age} & 1980\le\mathrm{Year}<2005\\ \beta_0+u_{0i}+(\beta_1+u_{1i})\mathrm{Y}_1+\beta_2\mathrm{Y}_1^2+\beta_3Y_2+\beta_4Y_3+\gamma_1\mathrm{SEX}+\gamma_2\mathrm{C\_Age} & \mathrm{Year}\ge2005\end{cases}$$

with

$$\Pr(Y=y_{ij}\mid\mu_{ij},\alpha)=\frac{\Gamma(y_{ij}+\alpha^{-1})}{\Gamma(\alpha^{-1})\Gamma(y_{ij}+1)}\left(\frac{1}{1+\alpha\mu_{ij}}\right)^{\alpha^{-1}}\left(\frac{\alpha\mu_{ij}}{1+\alpha\mu_{ij}}\right)^{y_{ij}}$$

$$\alpha=\frac{1}{\nu}$$

The residuals for the random intercept and linear slope of year are assumed to follow a multivariate normal distribution (MVN) with means, variances, and covariances specified by

$$\begin{bmatrix}u_{oi}\\u_{1i}\end{bmatrix}\sim MVN\left(\begin{bmatrix}0\\0\end{bmatrix},\ \begin{bmatrix}\sigma_0^2 & \sigma_{01}\\ \sigma_{01} & \sigma_1^2\end{bmatrix}\right).$$

RESULTS

Table 3 reports parameter estimates (standard errors), p-values, and goodness of fit statistics for Models A, B and C for hypertension. The fixed-effects parameter estimates are similar

among the three models for most predictors except sex. Results from Model C indicate that the prevalence rate of hypertension in 1970 was about 39.0% (exp(-3.245)*1000). The prevalence rate increased about 14.3% (exp(0.133+0.001)-1) in 1971. The linear increase rate of subsequent years increased about 0.1% (positive slope of quadratic term) every year before the coding transition in 1980. At the first few years of transition to new coding system in 1980, there was no increase in the prevalence rate (the linear increase rate for years from 1980 is significant lower than that for years before 1980). Similar increase in growth rate of prevalence was observed for the second and third phases. After another transition of coding system in 2005, the growth rate of prevalence of hypertension slowed down again (the linear increase rate for years from 2005 is significant lower than that for years before 2005). Table 3 also indicates that the prevalence of hypertension among females is higher than males, higher among older cohorts than younger cohorts. Results from Model C indicates that there were significant variations in both level and trend of prevalence of hypertension among different age by sex groups. Goodness of fit index reported in Table 3 shows that Model C (with random effect of intercept and slope) fits our data significantly better than both Model A and B (Vuong, 1989).

Parameters	**Model**		
	Model A	**Model B**	**Model C**
Fixed Effects			
Intercept	-3.139(0.05) ***	-3.129(0.082) ***	-3.245(0.097) ***
Sex	0.186(0.023) ***	0.185(0.094) ^	0.402(0.181) **
Age	0.049(0.001) ***	0.047(0.004) ***	0.046(0.005) ***
Year since 1970	0.132(0.008) ***	0.132(0.0062) ***	0.133(0.0059) ***
(Year since 1970)*(Year since 1970)	0.001(0.0003) ***	0.001(0.0002) ***	0.001(0.0002) ***
Year after 1980	-0.174(0.015) ***	-0.174(0.012) ***	-0.174(0.011) ***
Year After 2005	-0.024(0.013) ^	-0.023(0.01) **	-0.022(0.009) **
Scale parameter	0.077(0.005)	0.048(0.003)	0.037(0.002)
Variance Components			
Intercept	NA	0.030(0.012)	0.259(0.014)
Intercept & (Year since 1970)	NA	NA	-0.000(0.001)
Year since 1970	NA	NA	0.0001(0.00003)
Fit Statistics			
-2 Log Likelihood	10072	9842	9736
AIC	10088	9860	9758
BIC	10123	9866	9765

Notes: 1. ^ p <.10, * p <.05, ** p <.01, *** p <.001. Sex was coded 0 for male and 1 for female. Mean age for each group was centered at 50 years. NA: not applicable. 2. Model A: Conventional negative binomial regression; Model B: Model A plus a random intercept; Model C: Model A plus random intercept and random linear slope of time.

Table 3. Parameter Estimates (Standard Errors) from Models for Predicting Hypertension Prevalence (Logarithmic Transformation)

The overall prevalence of hypertension predicted by the three models is displayed in Figure 1 along with the observed prevalence. Model A over predicted the prevalence rate after 1990, while Models B and C appear to provide more accurate predictions than Model A.

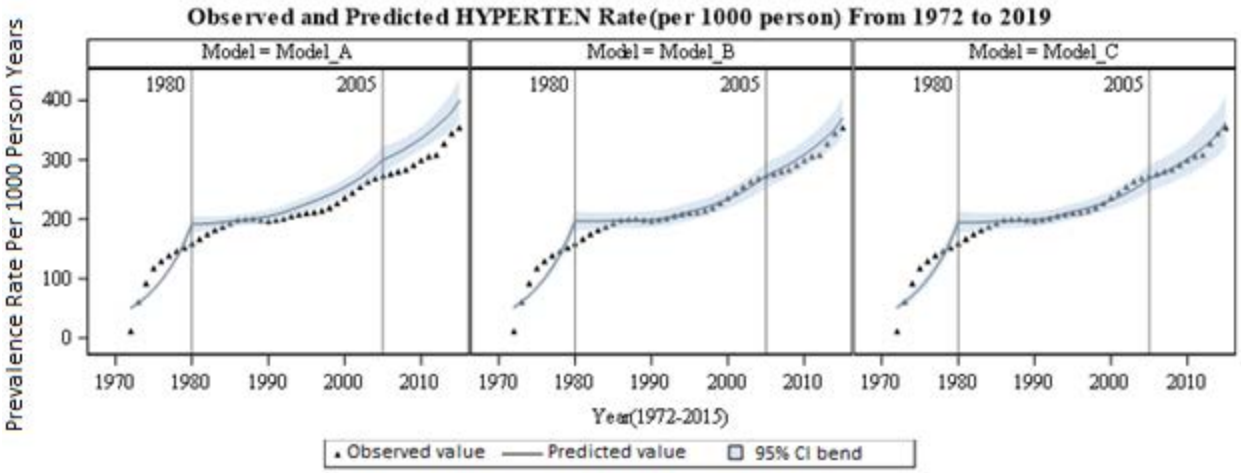

Figure 1. Observed and Predicted Prevalence Rate of Hypertension

The results from three models for COPD are shown in Table 4. The fixed effects parameter estimates are also quite similar among the three models for most all predictors. Results from Model C for COPD indicate that the prevalence rate of COPD in year of 1970 is about 5.6% (exp(-5.186)*1000). The prevalence rate increased about 13.5% (exp(0.129+0.002)-1) in 1971. The linear increase rate of subsequent years deceased about 0.2% (negative slope of quadratic term) every year before the coding transition in 1980. After the transition to a new coding system in 1980, the increase in the prevalence rate slowed down significantly (the linear increase rate for years from 1980 is significant lower than that for years before 1980). Similar increase in growth rate of prevalence was observed for the second and third phases. After another transition of coding systems in 2005, the growth rate of prevalence of COPD increased again (the linear increase rate for years from 2005 is significant higher than that for years before 2005). Table 4 also indicates that the prevalence of COPD among females is lower than males, higher among older cohorts than younger cohorts. Results from Model C indicate that there were significant variations in both level and trend of prevalence of COPD among different age by sex groups. Goodness of fit statistics reported in Table 4 shows that Model C (with random intercept and slope) fit data significantly better than both Model A and B.

Parameters	**Models**		
	Model A	**Model B**	**Model C**
Fixed Effects			
Intercept	-5.116(0.043) ***	-5.11(0.08) ***	-5.186(0.113) ***
Sex	-0.197(0.02) ***	-0.19(0.097) *	-0.23(0.153)
Age	0.076(0.001) ***	0.071(0.004) ***	0.076(0.004) ***
Year since 1970	0.121(0.007) ***	0.123(0.0049) ***	0.129(0.0044) ***
(Year since 1970)*(Year since 1970)	-0.002(0.0002) ***	-0.002(0.0002) ***	-0.002(0.0001) ***
Year after 1980	-0.053(0.013) ***	-0.054(0.009) ***	-0.057(0.006) ***
Year After 2005	0.037(0.011) ***	0.037(0.008) ***	0.037(0.005) ***
Scale parameter	0.056(0.004)	0.026(0.002)	0.008(0.001)
Variance Components			
Intercept	NA	0.032(0.013)	0.0749(0.038)

Intercept & (Year since 1970)	NA	NA	-0.002(0.001)
Year since 1970	NA	NA	0.0001(0.00005)
Fit Statistics			
-2 Log Likelihood	7411	7082	6620
AIC	7429	7100	6642
BIC	7465	7106	6649

Notes: 1. ^ p <.10, * p <.05, ** p <.01, *** p <.001. Sex was coded 0 for male and 1 for female. Mean age for each group was centered at 50 years. NA: not applicable. 2. Model A: Conventional negative binomial regression; Model B: Model A plus a random intercept; Model C: Model A plus random intercept and random linear slope of time.

Table 4. Parameter Estimates (Standard Errors) from Models for Predicting Log of Prevalence Rate of Chronic Obstructive Pulmonary Disease (COPD)

The overall prevalence of COPD predicted by the three models is displayed in Figure 2 along with observed prevalence. Model A over predicted the prevalence rate after 1995 and Model B over predicted the prevalence of COPD after 2005, while the Model C provided more accurate predictions than Models A and B.

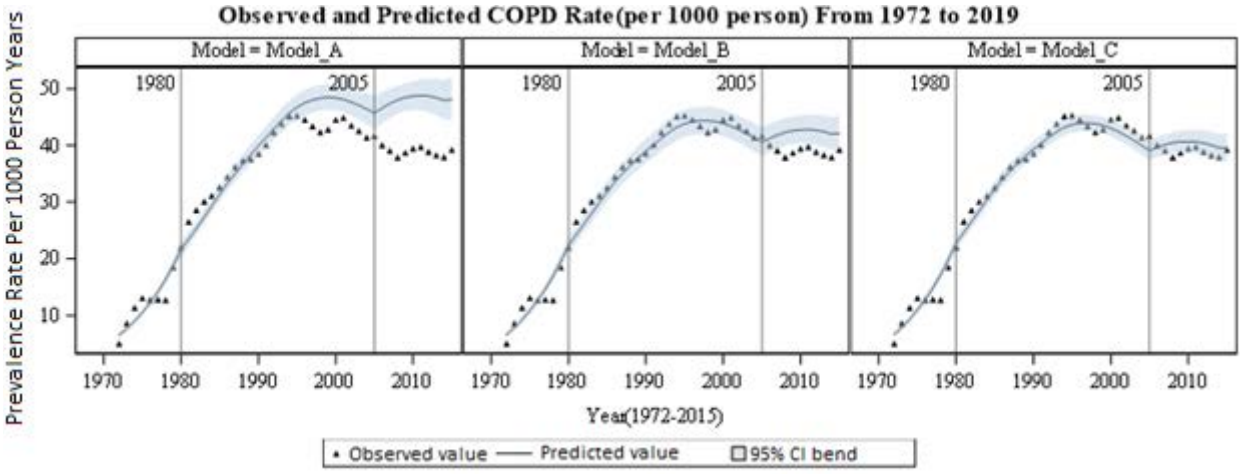

Figure 2. Observed and Predicted Prevalence Rate of Chronic Obstructive Pulmonary Disease (COPD)

DISCUSSION AND CONCLUSIONS

Public health researchers and decision makers are often interested in not only the disease prevalence rate, but also the trends over time and variations among different groups. The piecewise latent growth curve model with random effects can be used to provide more accurate estimations of trends in disease prevalence and inter-group differences in these trends. Also, it is very important to take into consideration the changes in disease coding while assessing the long-term trends in disease prevalence and its variation.

The results from the piecewise latent growth curve models show that the model with random effects of both level and trend provides better goodness of fit and the most accurate prediction of disease prevalence and long-term trends. Though the estimated average linear and quadratic slopes before the transition and the change after the transition in the random effects models were similar to those from fixed effect model, the random effect models take the group differences in the disease prevalence and trends into consideration, which makes the prediction more accurate, especially over long term period.

We compared several piecewise latent growth curve models. This included linear or nonlinear trend within each time segment before and after the transition points; and discontinuity with or without an elevation (i.e., change in slope) at the transition point and found the model with quadratic function of year and ICD impact on an immediate shift in slope was most appropriate for hypertension and COPD. However, this may not be true for other chronic diseases. A non-linear growth model (e.g., logistic curve model) might be needed for other diseases and there might be shifts in levels with the coding transition. The difference in growth rate before and after the transition might also need to be treated as a random effect. An increasingly complex model results in an increase in the number of random effects and nonlinear trends, we will need data at either individual level or aggregated data from large number of groups with longer time series.

The impact of changes in ICD coding might not take effect right at the starting year of a new coding system. In that case, the change point might need to be estimated from data. A piecewise latent growth curve model with unknown transition points is a type of nonlinear random effects model that is flexible to test the change point in segments of the regression model. However, estimation of the piecewise latent growth curve model with more random effects and nonlinear trends is computationally challenging.

REFERENCES

Bollen, K.A. & Curran, P.J. (2006). Latent curve models. A structural equation perspective. New York: Wiley.

Cudeck, R., & Klebe, K. J. (2002). Multiphase mixed-effects models for repeated measures data. Psychological Methods, 7, 41–63.

Dean, C. B. (1992). Testing for overdispersion in Poisson and binomial regression models. *Journal of the American Statistical Association 87*: 451–457.

Hilbe, J. M. (2011). Negative Binomial Regression. 2nd ed. Cambridge: Cambridge University Press.

Janssen, F.,& Kunst A. E. (2004). ICD coding changes and discontinuities in trends in cause-specific mortality in six Euopean countries, 1950-99. Bulletin of the World Health Organization, 82(12): 904-913.

Kohli N., & Harring J.R. (2013) Modeling growth in latent variables using a piecewise function. *Multivariate Behavior Research, 48*:370–97.

Leite W.L., & Stapleton L.M. (2011). Detecting growth shape misspecifications in latent growth models: an evaluation of fit indexes. *Journal Experimental Education, 79*:361–81.

Preacher K.J., Wichman A.L., MacCallum R.C., & Briggs N.E. (2008). Latent Growth Curve Modeling. Los Angeles, CA: Sage.

Singer J.D. & Willett J.B. (2003). “Applied Longitudinal Data Analysis”. New York: Oxford University Press.

Slavova, S., Costich, J. F., Luu, H., Fields, J., Gabella, B. A., Tarima, S., & Bunn, T. L. (2018). Interrupted time series design to evaluate the effect of the ICD-9-CM to ICD-10-CM coding transition on injury hospitalization trends. *Injury epidemiology, 5(1)*, 36.

Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica 57*: 307–333.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Dr. Lin Yan
University of Manitoba, Winnipeg, Canada
Lin.Yan@umanitoba.ca

ACKNOWLEDGEMENT

We acknowledge the Manitoba Centre for Health Policy for use of data contained in the Population Health Research Data Repository (HIPC #: 2017/2018/11; MCHP Project #: 2017-025). The results and conclusions are those of authors and no official endorsement by the Manitoba Centre for Health Policy, Manitoba Health, Healthy Living, and Seniors, or other data providers is intended or should be inferred, or other data providers is intended or should be inferred. Dr. Lisa M. Lix is supported by a Tier 1 Canada Research Chair.