Causes-of-Death Mortality: What Do We Know on Their Dependence?

Séverine Arnold (-Gaille)* and Michael Sherris†

Presented at the Living to 100 Symposium

Orlando, Fla.

January 8–10, 2014

* Faculty of Business and Economics, University of Lausanne, Switzerland, Severine.Gaille@unil.ch
† Australian School of Business, University of New South Wales, Australia, M.Sherris@unsw.edu.au

Causes-of-Death Mortality: What Do We Know on their Dependence?

Séverine Arnold (-Gaille)
Faculty of Business and Economics
University of Lausanne
Switzerland
Severine.Gaille@unil.ch

Michael Sherris
Australian School of Business
University of New South Wales
Australia
M.Sherris@unsw.edu.au

Abstract

Over the last century, the assumption usually made was that causes of death are independent, although it is well-known that dependencies exist. Recent developments in econometrics allow, through Vector Error Correction Models (VECM), to model multivariate dynamic systems including time dependency between economic variables. Common trends that exist between the variables may then be highlighted, the relation between these variables being represented by a long-run equilibrium relationship. In this work, VECM are developed for causes-of-death mortality. We analyze the five main causes of death across ten major countries representing a diversity of developed economies. The World Health Organization website provides cause-of-death information over about the last 60 years. Our analysis reveals that long-run equilibrium relationships exist between the five main causes of death, improving our understanding of the nature of dependence between these competing risks over recent years. It also highlights that countries had usually different past experience in regards to cause-of-death mortality trends and thus, applying results from one country to another may be misleading.

Keywords: Causes of death, mortality trends, VECM, dependence, common trends

Contents

List of Figures

List of Tables

1 INTRODUCTION

Models for trends in mortality rates for different ages and sexes as well as for different countries are often based on the assumption that past trends in historical data will continue in the future. Past mortality trends and variability reflect many factors and these include changes in the causes of death. These causes have differing age patterns and have shown different trends over recent years. At the same time, systematic changes in causes of death have often been shown or assumed common across the developing economies. Tuljapurkar et al. [2000] show how mortality declines have had common trends in the G7 countries although there is evidence of variability in those trends. Booth et al. [2006] also demonstrate common improvement trends based on the Lee-Carter model and variants of the model. Wilmoth [1995] shows how taking into account causes of death can influence projected trends and effectively highlights how cause-of-death trends are hidden in aggregate data.

Dependence between competing risks is important in constructing aggregate mortality rates. However, the relations that exist between the causes of death are not well understood. Usually an assumption is made that causes of death are independent. Cause elimination models as well as cause-delay models developed by Manton et al. [1980] and Jay Olshansky [1987] are two well-known examples. The independence assumption is also frequently implicit in cause-specific mortality forecasts, as mortality is usually projected for each cause independently and aggregated in the end to produce total mortality, see e.g. Tabeau et al. [1999], McNown and Rogers [1992] and Caselli et al. [2006].

Vector AutoRegressions (VAR) as well as Vector Error Correction Models (VECM) are tools developed in econometrics that give valuable information on the relations binding a set of variables, typically a set of economic variables. Indeed, beside including time dependency between the variables of interest and allowing for stochastic trends, these models use long-run equilibrium relationships through what is know as cointegration. These long-run equilibriums represent steady-states that exist between the variables under study. As a result, the application of these models to cause-of-death mortality rates will provide valuable information on their dependence and allow a better understanding

of the trends in cause-of-death mortality rates across countries.

In this paper, we estimate the common trends and relationships between the five main causes of death in ten countries. The paper shows that although some countries have similar trends in cause-of-death mortality rates, there are differences in groups of countries and in the form of the long-run common stochastic trends. Thus, it highlights that countries had usually different past experience in regards to cause-of-death mortality trends and that applying results from one country to another may be misleading. The resulting improved understanding of the relations existing between the causes of death can further be used in competing risk models and in constructing aggregate mortality rate trends. This will better inform estimates of future mortality trends and variability.

The paper begins with a brief description of VAR and VECM in Section 2. Section 3 summarizes the data source and cause-of-death mortality used to estimate the models. Results from the model fitting are then discussed in Section 4. Section 5 highlights implications for modeling mortality trends and concludes.

2 VAR AND VECM MODELS

Stationary variables, such as a vector of stationary cause-specific mortality rates, can be effectively modeled through Vector AutoRegressive (VAR) models. A pth-order vector autoregression, denoted as VAR(p), based on p lags of the variables in the model is written as

$$\mathbf{y}_t = \mathbf{c} + \mathbf{d}t + \mathbf{\Phi}_1\mathbf{y}_{t-1} + \mathbf{\Phi}_2\mathbf{y}_{t-2} + \cdots + \mathbf{\Phi}_p\mathbf{y}_{t-p} + \epsilon_t, \tag{1}$$

where the n variables at time t are denoted by the $(n \times 1)$ vector $\mathbf{y}_t$, $\mathbf{c}$ is a $(n \times 1)$ vector of constants, $\mathbf{d}$ is a $(n \times 1)$ vector of trends and $\mathbf{\Phi}_i$ is a $(n \times n)$ matrix of autoregressive coefficients for $i = 1, 2, \ldots, p$. The $(n \times 1)$ vector ϵ_t is a vector of white noise terms, with

$$E(\epsilon_t) = \mathbf{0}, \tag{2}$$

$$E(\epsilon_t\epsilon_l) = \begin{cases} \mathbf{\Omega} & \text{for } t = l \\ \mathbf{0} & \text{for } t \neq l, \end{cases} \tag{3}$$

where $\mathbf{\Omega}$ is a symmetric positive definite matrix. Hamilton [1994] and Lütkepohl [2005] are comprehensive references on these models.

However, variables are often non-stationary, also called integrated. The non-stationarity can be removed by differencing the variables if the process is integrated of order one, denoted $I(1)$. A VAR(p) can then be fitted to the differenced data. Nonetheless, non-stationary variables may have common stochastic trends. In such a case, the variables move together, influenced by the common trends. There might then exist a linear combination of the variables, such that the resulting relation is stationary, even if each variable is not. This stationary relation represents a long-run equilibrium relationship called cointegration. By differencing the data, any information about long-run trends present in the levels of the data is removed, such as a potential cointegrating relation. Therefore, fitting a VAR to the differenced data is not optimal. However, models that include cointegrating relations exist. They are called Vector Error Correction Models (VECM).

Formally, as described in Lütkepohl [2005] and following the notation used in Gaille and Sherris [2011], if the n variables in the vector $\mathbf{y}_t$ are all $I(1)$ then, if they are cointegrated, a long-run relationship given by

$$\beta_1 y_{1t} + \beta_2 y_{2t} + \cdots + \beta_n y_{nt} = 0$$

will hold on average in the long-run. Allowing for deviations from the long-run equilibrium relationship this becomes

$$\beta_1 y_{1t} + \beta_2 y_{2t} + \cdots + \beta_n y_{nt} = z_t, \tag{4}$$

where z_t is a stochastic variable representing that deviation. The variables are cointegrated if z_t is stationary, that is a long-run equilibrium exists.

The variables under study may be linked by more than one cointegrating relation, each relation being linearly independent from the others. The cointegrating relations are then represented in a matrix form, β, each column of that matrix being a cointegrating

relation, as in

$$\begin{aligned}
\beta'\mathbf{y}_t &= \mathbf{z_t}, && (5)\\
\beta &= (\beta_\mathbf{1} \quad \beta_\mathbf{2} \ldots \beta_\mathbf{r}),\\
&= \begin{pmatrix} \beta_{11} & \beta_{12} & \cdots & \beta_{1r} \\ \beta_{21} & \beta_{22} & \cdots & \beta_{2r} \\ \vdots & & & \vdots \\ \beta_{n1} & \beta_{n2} & \cdots & \beta_{nr} \end{pmatrix}, && (6)\\
\mathbf{y}_t &= (y_{1t} \quad y_{2t} \ldots y_{nt})',
\end{aligned}$$

where $\mathbf{z_t}$ is now a vector of r stochastic variables. The vector $\beta'\mathbf{y}_t$ is stationary and contains the r linearly independent cointegrating relations of the n variables in the process.[1] The columns of β are said to form a basis of the space of cointegration (Hamilton [1994]) when any other cointegrating relation can be expressed as a linear combination of $(\beta_\mathbf{1} \quad \beta_\mathbf{2} \ldots \beta_\mathbf{r})$. There are then exactly r cointegrating relations among the variables under study.

The cointegrating relations are incorporated in VAR modeling using an alternative VAR(p) representation (see, for example, Hamilton [1994] for a proof)

$$\nabla\mathbf{y}_t = \mathbf{c} + \xi_1\nabla\mathbf{y}_{t-1} + \xi_2\nabla\mathbf{y}_{t-2} + \cdots + \xi_{p-1}\nabla\mathbf{y}_{t-p+1} + \mathbf{\Pi}\mathbf{y}_{t-1} + \epsilon_t, \tag{7}$$

where

$$\begin{aligned}
\mathbf{\Pi} &= -(\mathbf{I}_n - \mathbf{\Phi_1} - \cdots - \mathbf{\Phi_p});\\
&= \alpha\beta';\\
&= \text{matrix of rank } r;\\
\alpha &= \text{a } (n \times r) \text{ loading matrix ;}\\
\beta &= \text{a } (n \times r) \text{ matrix containing the } r \text{ vectors}\\
&\quad \text{forming a basis of the space of cointegration;}
\end{aligned}$$

[1] Only variables integrated of order one are considered, and thus cointegrating relations between these variables are stationary. For a more general description, see Hamilton [1994] and Lütkepohl [2005].

$$\xi_i \quad = \quad -(\mathbf{\Phi_{i+1}} + \cdots + \mathbf{\Phi_p}) \quad \text{for } i = 1, \ldots, p-1.$$

Equation (7) is the Vector Error Correction Model of the cointegrated system. Since the first difference of an $I(1)$ process and the cointegrating relations are stationary, each element of Equation (7) is stationary. The loading matrix α measures the impacts cointegrating relations have on the variables under study. For example, the element α_{ij} measures the effect of the cointegrating relation j $(j = 1, \ldots, r)$ on the variable i $(i = 1, \ldots, n)$.

In order to estimate the cointegrating relations and the other parameters in Equation (7), we use Johansen's approach, which is the standard procedure when the number of cointegrating relations in the VECM is unknown. The steps to follow to estimate a VECM are described in Gaille and Sherris [2011]. In summary (Figure (1)):

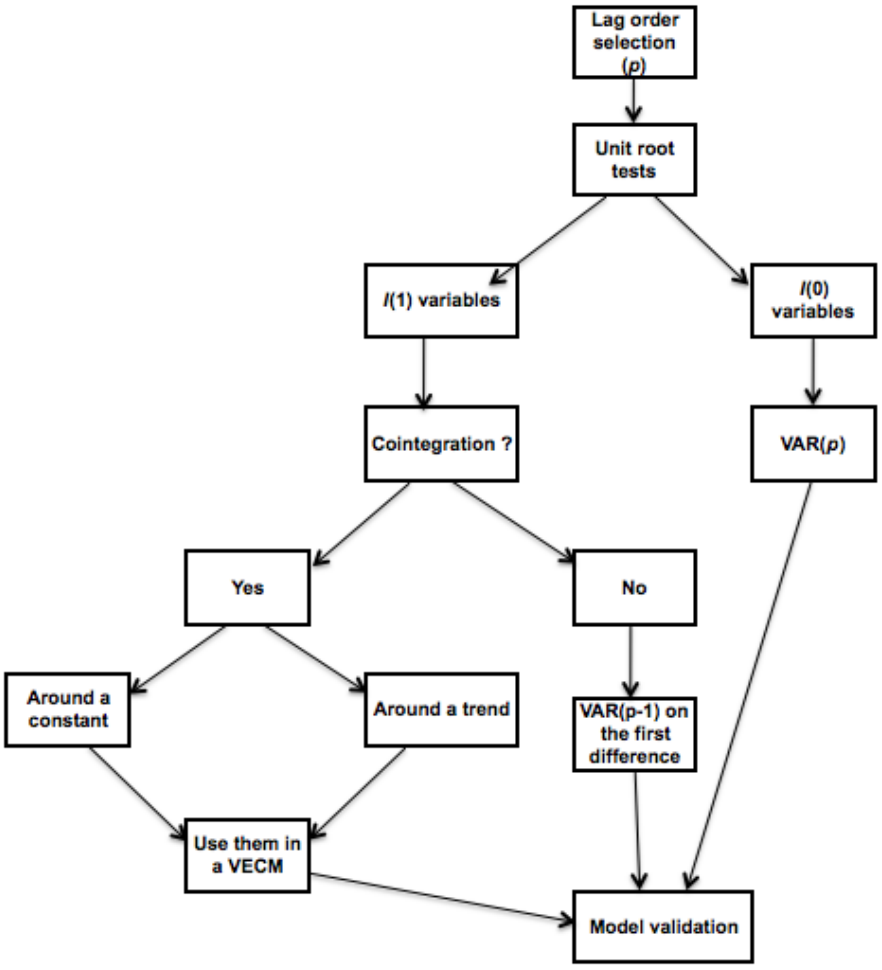


Figure 1: Steps to follow in a VECM analysis

This is an extension of the procedure described in Gaille and Sherris [2011].

1. The lag order of the VAR is selected through Akaike's Information Criteria (AIC), Hannan-Quinn Criterion (HQ), Schwarz Criterion (SC), Final Prediction Error (FPE).

2. Stationarity of the variables is considered through several unit root tests: the Kwiatkowski-Phillips-Schmidt-Shin test (KPSS), the Augmented Dickey-Fuller test (ADF), the Phillips-Perron test (PP) or the Elliot-Rothenberg-Stock test (ERS).

3. Fitting a VAR(p) is appropriate if all the variables are stationary. A stationary variable is also denoted $I(0)$. However, the trace test and the maximum-eigenvalue test of the Johansen's approach are used in order to find the number of cointegrating relations when some of the variables are $I(1)$. Besides, as mentioned in Gaille and Sherris [2013], the cointegrating relations may be stationary around a constant or a trend. This can also be tested through the Johansen's approach, while testing for the number of cointegrating relations.

4. A VAR($p-1$) on the first difference is estimated when the variables are $I(1)$ and not cointegrated. Otherwise, the appropriate VECM should be found.

5. The residuals of the model are tested for normality and autocorrelations.

3 DATA

Central death rates are determined as the number of persons for each age group, sex, and country who die in a particular year of a specific cause, divided by the mid-year population. Data were obtained from the Mortality Database (World Health Organization [2012]) administered by the World Health Organization (WHO) which contains demographic information, including the number of deaths by cause of death, for various countries over the last 50 or 60 years. The data are generally divided into five-year age-groups.

The five developed countries with the highest population are studied along with five other countries selected such that a diversity in population size is also represented in our analysis. Although Germany is the third developed country with the highest population,

it is not included in our study since the population and death numbers are, before 1990, split between the Former Democratic Republic, the Former Federal Republic and West Berlin and since the Former Democratic Republic has no available data for the period preceding 1969. Besides, developing countries are not included, their data being less reliable. The ten countries studied are, by decreasing population size, USA (1950–2007), Japan (1950–2009), France (1952–2008), Italy (1951–2003), England and Wales (1950–2009), Australia (1950–2004), Sweden (1951–2010), Switzerland (1951–2007), Singapore (1955–2009) and Norway (1951–2009).

The International Classification of Diseases (ICD) is an internationally recognized classification of the causes of death aiming at ensuring consistency between countries (Table (1a)). Under the ICD, the underlying cause of death is specified as *the disease or injury which initiated the train of morbid events leading directly to death, or the circumstances of the accident or violence which produced the fatal injury*. We consider the five main ICD causes, which are: diseases of the circulatory system, cancer, diseases of the respiratory system, external causes, and infectious and parasitic diseases. Gaille and Sherris [2013] noticed that these major causes accounted for more than 80% of deaths in recent years, and made up approximately 60% – 70% of deaths 50 years ago.

The same database as in Gaille and Sherris [2011, 2013] is used in this paper. Therefore, in order to analyze data consistently over time and across countries, some adjustments are made, such as distributing the number of deaths of unknown age proportionally across the age range, grouping ages 85 and above as well as ages one to four, and taking into account the changes of classification over time. Since the ICD changed three times between 1950 and 2010, from ICD-7 to ICD-10 (Table (1b) presents the dates at which the countries adopted new classifications), the raw data are not directly comparable over time. Figure 2 introduces death rates in the USA for different causes of death. The ICD change in 1968 highly impacts the diseases of the circulatory system, while the ICD change in 1999 clearly influences the infectious & parasitic diseases, as noticed with the upward jumps. To get data that are comparable over time, comparability ratios are used as described in Gaille and Sherris [2011]. Discontinuities in the death rates at the dates introduced in Table (1b) are then removed using these comparability ratios.

Table 1: International Classification of Diseases

(a) Coding system

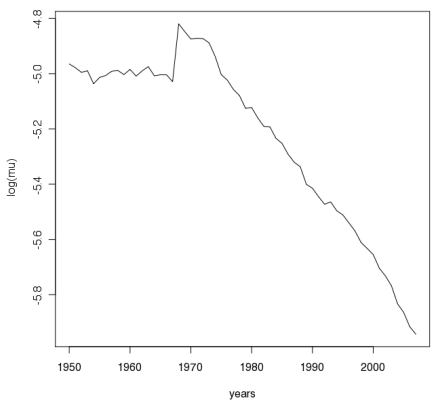
Causes of death	ICD 7		ICD 8
	Singapore	Other countries	
Circulatory system	B024-B029	A079-A086	A080-A088
Cancer	B018-B019	A044-A060	A045-A061
Respiratory system	B030-B032	A087-A097	A089-A096
External causes	B047-B050	A138-A150	A138-A150
Infectious and parasitic diseases	B001-B017	A001-A043	A001-A044

Causes of death	ICD 9	ICD 10	
		Switzerland	Other countries
Circulatory system	B25-B30	1064	I00-I99
Cancer	B08-B17	1026	C00-D48
Respiratory system	B31-B32	1072	J00-J99
External causes	B47-B56	1095	V00-Y89
Infectious and parasitic diseases	B01-B07	1001	A00-B99

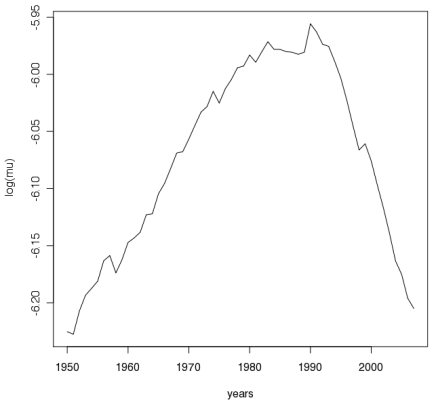
(b) Adoption of new classifications

Country	ICD change	Year
USA	ICD7-8	1968
	ICD8-9	1979
	ICD9-10	1999
Japan	ICD7-8	1968
	ICD8-9	1979
	ICD9-10	1995
France	ICD7-8	1968
	ICD8-9	1979
	ICD9-10	2000
Italy	ICD7-8	1968
	ICD8-9	1979
England and Wales	ICD7-8	1968
	ICD8-9	1979
	ICD9-10	2001

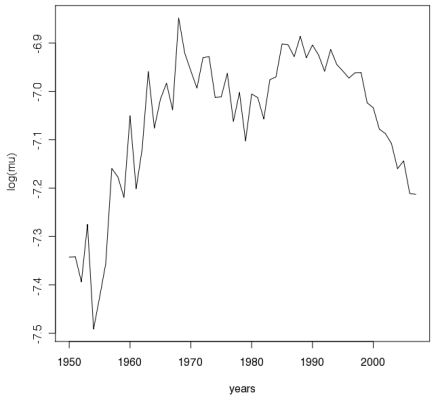
Country	ICD change	Year
Australia	ICD7-8	1968
	ICD8-9	1979
	ICD9-10	1998
Sweden	ICD7-8	1969
	ICD8-9	1987
	ICD9-10	1997
Switzerland	ICD7-8	1969
	ICD8-10	1995
Singapore	ICD7-8	1969
	ICD8-9	1979
Norway	ICD7-8	1969
	ICD8-9	1986
	ICD9-10	1996



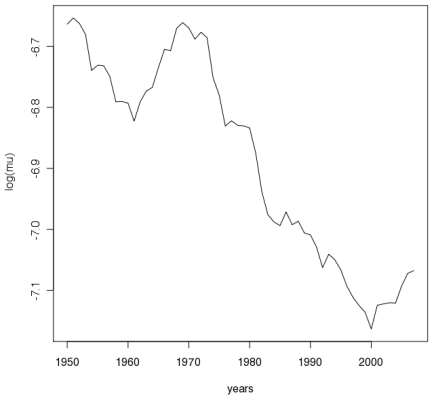
(a) Circulatory system



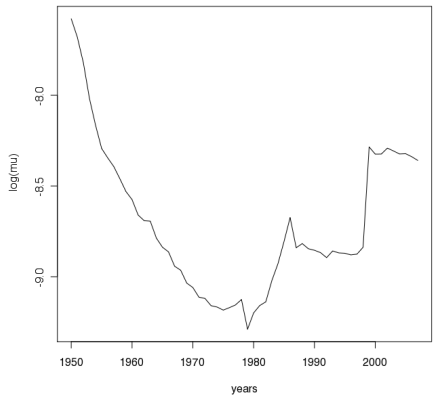
(b) Cancer



(c) Respiratory system

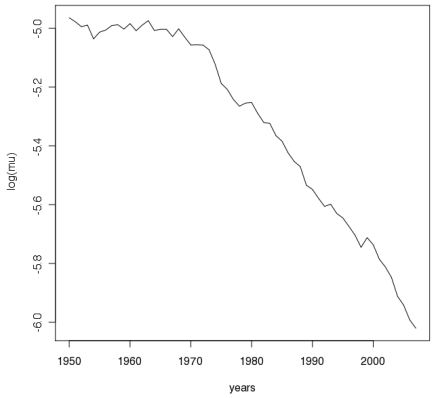


(d) External causes

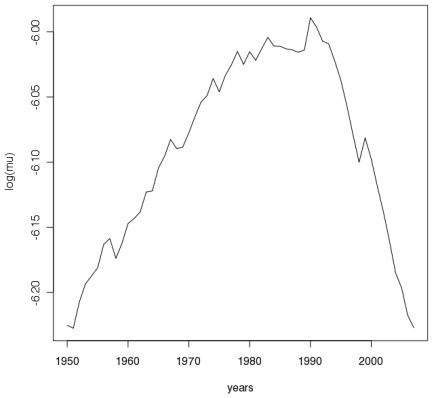


(e) Infectious & parasitic diseases

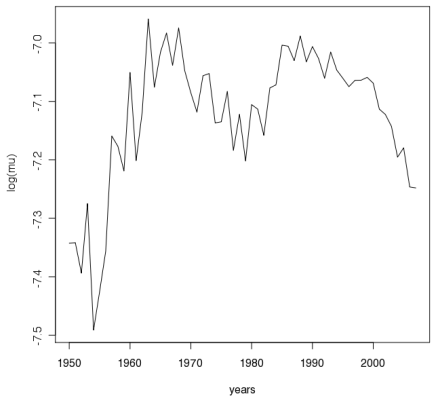
Figure 2: Observed cause-specific log-death rates, males in USA



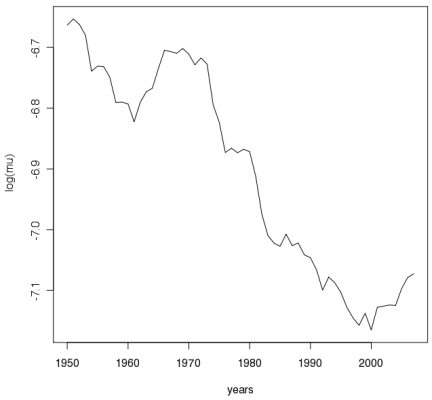
(a) Circulatory system



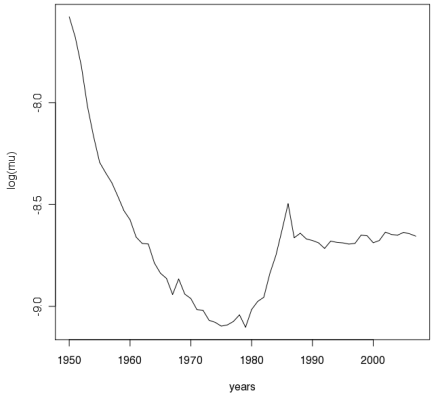
(b) Cancer



(c) Respiratory system



(d) External causes



(e) Infectious & parasitic diseases

Figure 3: Cause-specific log-death rates smoothed by the comparability ratios, males in USA

Figure 3 presents the resulting graphs for the USA. The two jumps noticed in Figure 2 are removed. The analysis in this paper is applied to these adjusted death rates.

Finally, trends by cause of death are examined using an age-standardized country-specific central death rate. To allow for changes in the age structure of the population, the aggregate country-specific death rate is denoted by $m^*_{c,t,d,s}$, where

$$\begin{aligned} m^*_{c,t,d,s} &= d^*_{c,t,d,s}/l_{c,LY_c,s}, \\ d^*_{c,t,d,s} &= \sum_x (m_{x,c,t,d,s} \times l_{x,c,LY_c,s}), \end{aligned} \tag{8}$$

and

$$\begin{aligned} m_{x,c,t,d,s} &= \text{central death rate of country } c, \text{ at time } t, \text{ for cause of death } d, \\ &\quad \text{and for a person of sex } s, \text{ and age } x; \\ l_{x,c,LY_c,s} &= \text{mid-year population of sex } s, \text{ age } x, \text{ in country } c \text{ and year } LY_c; \\ l_{c,LY_c,s} &= \sum_x (l_{x,c,LY_c,s}); \\ &= \text{mid-year population of sex } s, \text{ in country } c \text{ and year } LY_c; \\ LY_c &= \text{last year under observation for country } c. \end{aligned}$$

The population of the last year under observation is used as a base. Total number of deaths in a particular year t is determined as if the mid-year population of that year was the same as the population of the last year of the data period. Therefore, $m^*_{c,t,d,s}$ refers to the country cause-specific death rate in year t, assuming that the population is constant over the complete period under observation and fixed at the level of the last observed year.

4 LONG-RUN TRENDS FOR CAUSES OF DEATH

The VECM is estimated across the ten major countries for males and females, using Johansen's procedure. Long-run equilibrium relationships are estimated between the five main causes of death. The analysis is applied to the logarithm of $m^*_{c,t,d,s}$. Since 20 different models are estimated (ten countries, two genders), more than 300 tests were

performed. Therefore, only summaries of most important results are presented for ease of presentation and to highlight the key findings of the paper. Details are available from the authors upon request.

4.1 Lag Order Selection

Out of the four tests performed, a lag order of one or two is indicated as optimal. A VAR(1) is the most suitable model for the aggregate standardized log-death rates for causes of death in each of the ten analyzed countries, except for females in France and Switzerland as well as males in Australia, see Table 2.

Table 2: Number of past values to take into consideration in a VECM analysis

	Males	Females
USA	1	1
Japan	1	1
France	1	2
E&W	1	1
Italy	1	1
Australia	2	1
Sweden	1	1
Switzerland	1	2
Singapore	1	1
Norway	1	1

4.2 Unit Root Tests

KPSS, ADF, PP and ERS tests are performed on the data. A cause of death is said stationary when at least three out of the four tests accept it at a five percent significance level. When some doubts still remain, several models are tested and the one with non-autocorrelated and normally distributed residuals (model validation criteria) is chosen. Table (3) summarizes the causes of death that are stationary according to these tests.

Table 3: Stationarity of the five main causes of death

	Males	Females
USA	All causes: UR	All causes: UR
Japan	All causes: UR	All causes: UR
France	All causes: UR	All causes: UR
E&W	All causes: UR	Respiratory: S Other causes: UR
Italy	Respiratory: S Other causes: UR	Respiratory: S Other causes: UR
Australia	All causes: UR	All causes: UR
Sweden	All causes: UR	All causes: UR
Switzerland	Respiratory: S Other causes: UR	All causes: UR
Singapore	I&P: S Other causes: UR	I&P: S Other causes: UR
Norway	Respiratory: S Other causes: UR	(Cancer), Respiratory: S Other causes: UR

UR = Unit root, that is a non-stationary variable
S = Stationary variable
I&P = Infectious and parasitic diseases.
This table describes the stationarity of the log-death rate $\log m^*_{c,t,d,s}$.

Across the countries, most of the cause-of-death log-death rates show evidence of non-stationarity and have stochastic trends. The major exceptions are the diseases of the respiratory system. In the United States, Japan, France, England and Wales (males), Australia, Sweden and Switzerland (females), the five main causes of death are non-stationary. In England and Wales (females), Italy, Switzerland (males) and Norway, log-death rate for diseases of the respiratory system is the only rate that is stationary. Singapore is different with log-death rate for infectious and parasitic diseases as the only stationary cause of death. This is expected to express the climate of this country, as Singapore is the only country with a tropical weather.

4.3 Long-Run Equilibrium Relationships

The number of estimated cointegrating relations is summarized in Table (4), based on the trace and maximum-eigenvalue tests of the Johansen's procedure. These two tests assess the number of long-run equilibrium relationships among the causes of death. Several model assumptions are tested and the most efficient one according to the model validation criteria is shown.

Table 4: Number of cointegrating relations among the five main causes of death

	Males	Females
USA	1	1
Japan	1	1
France	1	1
E&W	1	1
Italy	1	1
Australia	1	1
Sweden	1	1
Switzerland	2	2
Singapore	2	2
Norway	2	1

In general, there is at least one cointegrating relationship between the causes of death in each country showing that these rates have changed with common stochastic trends. These long-run equilibrium relationships determine how changes in causes of death move relative to each other.

Besides, tests performed through the Johansen's procedure indicate that a trend should be included in these cointegrating relations, except for females in Switzerland and Norway. Thus, the long-run equilibrium relationship is stationary around a trend meaning that the cause-of-death death-rates are either getting closer to each other over time or moving apart.

4.4 Fitted VECM for Causes of Death

Since the main interest in this work is to get a better understanding of the dependence that exists between causes of death, only cointegrating relations are presented and discussed. Indeed, these relations provide insights into the long-run relationships that exist between the variables of interest. However, details of each VECM are available from the authors upon request.

Table 5 describes the cointegrating relations found for the ten countries and both genders. These relationships reflect the historical data and the relative past changes in cause-specific mortality. It indicates that even if death rates were stochastically changing over time, the relations existing between these rates were constant. For example, for males in United States, decreases (increases) in the log-mortality rate of the circulatory system are associated with either increases (decreases) in log-death rate for cancer, or decreases (increases) in log-death rate for respiratory diseases, external causes of death or the infectious and parasitic diseases, or a combination of these impacts, so that the overall change is stationary. Thus, it provides a better understanding of the dependence existing between the causes of death.

Some similarities are found in the cointegrating relations across countries. For example, males in Sweden, in England and Wales as well as females in Sweden show similar relative changes. The five causes of death have a coefficient with the same sign. In these three cases, a decrease in mortality due to cancer was associated, in the past, with an

Table 5: Cointegrating relations between the five main causes of death

		Males					Females				
		Circulatory system	Cancer	Respiratory system	External causes	I&P	Circulatory system	Cancer	Respiratory system	External causes	I&P
USA	1st	10.69456	5.59757	-8.89295	-7.67594	-1.27201	-12.33722	21.23366	3.86850	9.11957	-2.44770
Japan	1st	7.88759	-18.19221	7.83418	1.91276	-5.31981	-7.04107	34.80843	-6.29159	-2.51984	2.79203
France	1st	3.61051	5.58839	9.92957	-8.57915	-9.35086	1.54265	-126.96155	13.87866	19.19450	-3.03437
E&W	1st	17.07335	16.15817	1.15538	19.21024	9.13450	23.66970	-41.64988	-8.28387	-0.97493	0.10822
Italy	1st	-4.65727	-15.23929	13.68104	10.70673	-2.51913	2.06643	7.74975	-11.33527	-0.84202	2.69301
Australia	1st	18.19951	-29.56920	-21.50232	10.31447	-1.79548	-6.94275	-13.93179	-9.76149	7.57787	3.14756
Sweden	1st	-10.08487	-3.35171	-5.54287	-6.00030	-7.55106	-23.67052	-3.53851	-2.54732	-1.99079	-4.38972
Switzerland	1st	-5.28092	9.94069	7.37395	-6.42822	-1.45477	-10.32052	15.33513	-1.79622	9.36036	1.06028
	2nd	-20.49919	28.31404	1.66893	5.96531	-1.34666	4.71468	-15.33328	-7.44481	0.59118	4.88562
Singapore	1st	15.18999	-8.63029	-5.62373	-4.32227	3.30489	-2.56778	8.37707	-4.68442	4.53691	-3.85025
	2nd	-6.06163	-0.48202	5.27013	-3.58569	3.53102	-12.32402	9.30955	4.75597	4.87606	0.64257
Norway	1st	12.75705	-8.84768	-13.00349	-7.90789	2.36239	10.63062	-16.28402	-11.00310	-9.68837	-0.54553
	2nd	15.80298	-4.81822	2.46397	10.22583	7.74822	-	-	-	-	-

I&P = Infectious and parasitic diseases.
These results show, for example, that the VECM for males in the United States has an estimated long-run equilibrium relationship given by

$$10.69 \times CircSyst_t + 5.60 \times Cancer_t - 8.89 \times RespSyst_t - 7.68 \times ExtCauses_t - 1.27 \times I\&P_t = z_t,$$

where z_t is a stationary variable.

increase in log-death rates of either or a combination of the four remaining causes.

However, these similarities are rare. Few countries have a similar cointegrating relation and thus, significant variations in trends between these cause-of-death rates are found across the ten countries. Concretely, it means that a decrease in mortality due to cancer, for example, will have different impacts on the other causes of death across countries, if past observed relations are assumed to hold in future. It can then be misleading to use the experience of an apparently similar country in order to model the mortality of another country of interest.

4.5 Model Validation

The residuals of the model are tested for normality with three tests (for details, see Lütkepohl [2005]): First a test based on the skewness of a normal distribution; second a test based on the kurtosis of a normal distribution; third, a combination of the first two tests, labeled *both* in Tables (6) and (7). Any remaining autocorrelations among the residuals are tested with the Portmanteau statistic adjusted for small samples (as in Lütkepohl [2005]). The null hypothesis is that there is no autocorrelation among the residuals up to 15 and 25 lags. Tables (6) and (7) summarize the significance of the tests for males and females respectively.

The null hypothesis of no-autocorrelation up to 15 or 25 lags as well as the null hypothesis of normality are, in most cases, accepted at a five percent significance level. For males in Italy as well as females in Singapore and England and Wales, the null hypothesis of normality is rejected. Despite this, the estimated VECM capture the trends in the cause-of-death data and provide a good fit based on the model assumptions.

5 DISCUSSION AND CONCLUSION

This paper presents the first results of an international comparison of potential long-run equilibrium relations between cause-specific mortality rates. By considering aggregate cause-of-death mortality rates and using models with long-run common stochastic trends, it is possible to estimate equilibrium relationships arising from different causes of

Table 6: Tests on residuals of the fitted VECM, males

	Portmanteau test		Normality tests		
	15 lags	25 lags	skewness	kurtosis	both
USA	***	***	***	*	***
Japan	***	***	***	***	***
France	***	***	***	***	***
E&W	***	***	***	***	***
Italy	***	***	-	-	-
Australia	***	***	***	-	**
Sweden	***	***	***	***	***
Switzerland	***	*	***	***	***
Singapore	***	***	***	**	***
Norway	***	***	***	***	***

* The null hypothesis is accepted at a one percent significance level.
** The null hypothesis is accepted at a 2.5% significance level.
*** The null hypothesis is accepted at a five percent significance level.
– The null hypothesis is rejected.

Table 7: Tests on residuals of the fitted VECM, females

	Portmanteau test		Normality tests		
	15 lags	25 lags	skewness	kurtosis	both
USA	***	***	***	***	***
Japan	-	***	***	***	***
France	**	***	***	***	***
E&W	***	***	-	-	-
Italy	***	***	***	*	**
Australia	**	***	***	***	***
Sweden	***	***	***	***	***
Switzerland	***	***	***	**	***
Singapore	***	***	-	-	-
Norway	**	***	***	***	***

* The null hypothesis is accepted at a one percent significance level.
** The null hypothesis is accepted at a 2.5% significance level.
*** The null hypothesis is accepted at a five percent significance level.
– The null hypothesis is rejected.

death. Comparing these trends across countries allows to identify countries with similar trends. This study uses a multivariate dynamic system to model log-death rates for the five main causes of death across ten countries. VECM are found to fit accurately the historical data and the dynamics of cause-specific death rates. Two important conclusions are drawn from these results.

First, long-run equilibrium relationships exist between the mortality rates for the five main causes of death in every country. This confirms the nature of dependence between these competing risks. The often made assumption of independence between mortality rates for causes of death is shown not to hold as these rates have common stochastic trends at a country level. Long-run equilibrium relationships should not be disregarded in any analysis considering the causes of death and should be included in new forecasting mortality models.

Second, the study demonstrates that countries tend to have different past experience, even if some groups of countries have a similar one. Causes of death have shown differing patterns of improvement and these patterns tend to vary across countries. If past trends are believed to be maintained in the future, a shock in some cause-specific mortality rates, e.g. a cure for cancer, will not have the same impact across countries. Thus, using the experience of some other and apparently similar countries in order to improve the model of some country-cause-specific mortality rates might be dangerous and misleading.

This paper presents new tools (cointegration and VECM) for modeling the dependence between causes of death and thus, offers new perspectives in such analyses. Preliminary results were presented but there is a wide range of potential additional developments. For example, this paper only analyses age-standardized death rates, since applying a VECM to age and cause-specific death rates would result in a model with far too many parameters. Further studies are needed in order to find the optimal way to model age and cause-specific mortality in a single model. Taking these new relations into account in the modeling and forecasting process will greatly improve the analysis of cause-specific mortality rates, as recently demonstrated with age-standardized death rates for Swiss females by Gaille and Sherris [2013].

References

H. Booth, Rob J. Hyndman, L. Tickle, and P. de Jong. Lee-Carter mortality forecasting: a multi-country comparison of variants and extensions. *Demographic Research*, 15: 289–310, 2006.

Graziella Caselli, Jacques Vallin, and Marco Marsili. How useful are the causes of death when extrapolating mortality trends. An update. *Social Insurance Studies from the Swedish Social Insurance*, 4, 2006.

S. Gaille and M. Sherris. Modeling Mortality with Common Stochastic Long-Run Trends. *The Geneva Papers on Risk and Insurance - Issues and Practice*, 36(4):595–621, 2011.

S. Gaille and M. Sherris. Forecasting Mortality Trends allowing for Cause-of-Death Mortality Dependence. *North American Actuarial Journal*, 17(4):273–282, 2013.

J. D. Hamilton. *Time Series Analysis*. Princeton University Press, 1994.

S. Jay Olshansky. Simultaneous/Multiple Cause-Delay (SIMCAD): An Epidemiological Approach to Projecting Mortality. *Journal of Gerontology*, 42(4):358–365, 1987.

H. Lütkepohl. *New Introduction to Multiple Time Series Analysis*. Crown Publishing Group, 2005.

Kenneth G. Manton, Clifford H. Patrick, and Eric Stallard. Mortality Model Based on Delays in Progression of Chronic Diseases: Alternative To Cause Elimination Model. *Public Health Reports*, 95:580–588, 1980.

R. McNown and A. Rogers. Forecasting cause-specific mortality using time series methods. *International Journal of Forecasting*, 8:413–432, 1992.

E. Tabeau, P. Ekamper, C. Huisman, and A. Bosch. Improving Overall Mortality Forecasts by Analysing Cause-of-Death, Period and Cohort Effects in Trends. *European Journal of Population*, 15:153–183, 1999.

S. Tuljapurkar, Nan Li, and C. Boe. A universal pattern of mortality decline in the G7 countries. *Nature*, 405:789–792, 2000.

J. R. Wilmoth. Are mortality projections always more pessimistic when disaggregated by cause of death? *Mathematical Population Studies*, 5(4):293–319, 1995.

World Health Organization. WHO Mortality Database, April 2012. `http://www.who.int/whosis/mort/download/en/index.html`.