Survival Analysis with Multiple Causes of Death

Extending the Competing Risks Model

Margarita Moreno-Betancur,[a,b,c] *Hamza Sadaoui,*[c] *Clara Piffaretti,*[c,d] *and Grégoire Rey*[c]

Abstract: Statistics on mortality related to each disease are usually based on the so-called underlying cause of death, which is selected from the diseases declared on the standardized death certificate using international rules. However, the assumption that each death is caused by exactly one disease is debatable, particularly with an aging population in an era where infectious diseases are replaced by chronic and degenerative diseases. The need to consider multiple causes of death has been acknowledged in epidemiologic research, with a growing body of literature producing statistics based on any mention of a disease on the death certificate. Yet there has not been a formal framework proposed for the statistical modeling of death arising from multiple causes. We propose a model for multiple cause of death data grounded on an empirical approach that assigns weights to each cause on the death certificate. We describe how this model for multiple-cause mortality, which extends the usual competing risks model used to conceptualize single-cause mortality, can serve to study the burden and etiology of mortality related to each disease, particularly using Cox regression methodology. We discuss how the multiple-cause, single-cause, and "any-mention" approaches compare in this regard. A simulation study and an application to a study of socioeconomic inequalities in mortality show the value of the proposed methods for exploiting this precious source of data to gain new insights, especially for certain diseases. See video abstract at, http://links.lww.com/EDE/B84.

(*Epidemiology* 2017;28: 12–19)

Submitted 22 July 2015; accepted 21 June 2016.

From the [a]Clinical Epidemiology and Biostatistics Unit, Murdoch Childrens Research Institute, Melbourne, Australia; [b]Department of Epidemiology and Preventive Medicine, Monash University, Melbourne, Australia; [c]Inserm CépiDc, Epidemiology Centre on Medical Causes of Death, Le Kremlin-Bicêtre, France; and [d]AP-HP, Assistance publique-Hôpitaux de Paris, Paris, France.

Supported by a Centre of Research Excellence grant from the Australian National Health & Medical Research Council, ID#1035261, to the Victorian Centre for Biostatistics (ViCBiostat), and a research grant from Institut National du Cancer (INCa, 2012-1-PL SHS-05-INSERM 11-1) and Cancéropôle Ile-de-France.

The authors report no conflicts of interest.

SDC Supplemental digital content is available through direct URL citations in the HTML and PDF versions of this article (www.epidem.com).

Correspondence: Margarita Moreno-Betancur, Clinical Epidemiology and Biostatistics Unit, Murdoch Childrens Research Institute, 50 Flemington Road, Parkville, Victoria 3052, Australia. E-mail: margarita.moreno@mcri.edu.au.

ISSN: 1044-3983/16/2801-0012
DOI: 10.1097/EDE.0000000000000531

The importance of cause of death data for health research is reflected in longstanding international efforts to standardize death certificates and disease coding through the World Health Organization's (WHO) International Classification of Diseases and Related Health Problems, now in its 10th Revision (ICD-10). The international form of the death certificate has two parts, as shown in the schematic figure provided in Volume 2 of the ICD-10 (page 24).[1] In part I, the physician should describe the primary morbid process leading to death, with the immediate cause on the first line and the so-called underlying cause of death on the last line. In part II, any other diseases that contributed to the death are reported. Certificates are not always correctly filled, which is why there is a set of WHO rules to select the underlying cause.[1,2]

Statistics on mortality related to each disease, henceforth referred to as disease-related mortality, are primarily based on the underlying cause of death. However, the assumption that all deaths are caused by only one disease is debatable, particularly with an aging population in an era where infectious diseases are replaced by chronic and degenerative diseases, several of which may be present at the moment of death.[3] Thus, there has been an increased use of so-called multiple cause of death approaches, a term designing any approach that examines other diseases reported on the death certificate.[4] An example is the recalculation of mortality rates attributed to a disease by considering any mention of it on the certificate.[5]

Yet, there has not been a formal framework proposed for the statistical modeling of mortality that acknowledges that death may be caused by several disease processes acting concurrently. We propose a model for multiple cause of death data by building on the weighting approaches described here and elsewhere.[6] We discuss how this model can be used to study disease-related mortality and compare it with other approaches, using simulations and a study of socioeconomic inequalities for illustration.

MODELING FRAMEWORK

Goals for Disease-related Mortality Statistics

The following two goals are of high relevance for health research:

Goal 1: To establish public health priorities, it is essential to quantify, in absolute terms, the burden of mortality attributable to each disease.

Goal 2: To understand the epidemiology of disease-related mortality, it is necessary to quantify the effect of certain factors (i.e., exposures, risk factors, etc.) on the force of mortality driven by each disease process, whether other separate disease processes are simultaneously present or not.

To address goals 1 and 2 using death certificate data, it has been emphasized that the method used should (1) acknowledge the multiple diseases that contributed to the death, as declared on the certificate; (2) count each death only once, regardless of how many diseases are reported on the certificate; and (3) reflect the relative importance of each disease in the occurrence of death compared with other contributing causes.[7,8] Next we discuss two common approaches to modeling these data and their limitations regarding these requirements.

Single-cause Model and Any-mention Approach

The single-cause model of mortality attributes each death entirely to one disease, taken to be the underlying cause of death from the WHO rules with death certificate data. This model thus violates requirements (1) and (3). The study of mortality related to a disease of interest is based here on the multistate model in Figure 1, which is the standard "competing risks" model from the survival analysis literature.[9] All individuals begin at the "alive" state and, when they die, move on to exactly one of two absorbing states representing, respectively, death with the disease of interest (state 1) or another disease (state 0) selected as underlying cause. The outcome for each individual is bivariate, consisting of the time-to-death T and the binary state indicator. The goals above are usually addressed by modeling the so-called cause-specific hazards, denoted by $\tilde{\lambda}_k(t)$ at time $t > 0$ for state $k \in \{0,1\}$, which are the instantaneous rates of transitions into each of the states. Actually, the epidemiologic cause-specific mortality rate approximates the mean cause-specific hazard over the follow-up period. The so-called cumulative incidence function is used for other purposes (see eAppendix, Section 1, http://links.lww.com/EDE/B77).

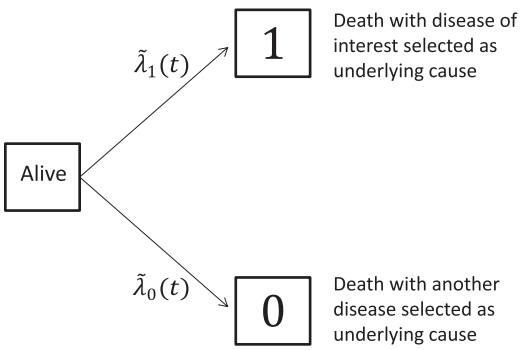

FIGURE 1. Multistate model for studying mortality related to the disease of interest based on the single-cause approach.

The "any-mention" approach considers any mention of the disease of interest on a certificate as an event, and is currently used in the context of recalculating disease-related mortality rates. This approach violates requirements (2) and (3). Each deceased individual contributes to death counts for as many diseases as appear on the certificate. Thus, the statistical units are items in death certificates instead of deaths, and this complicates the interpretation of estimates and gives the illusion of increased statistical power. Furthermore, all diseases mentioned on the certificate are assumed to have contributed equally to the death.

Multiple-cause Model

The proposed methodology is based on attributing a positive weight to each disease on the certificate such that the sum of the weights for all diseases mentioned is one. The aim of this procedure is to allow for all contributing diseases to be represented in the analysis (requirement (1)) but in such a way that each death has an equal influence in the analysis regardless of the number of diseases mentioned (requirement (2)). Ideally, the weight attributed to each disease should reflect the importance of its role in causing that death relative to other diseases (requirement (3)). The notion of a set of weights with these properties is a convenient way to address all these requirements and exploit all the information in death certificates, but it represents only an approximation to the actual causal process through which diseases act together to cause death (see "Discussion"). In particular, regarding the last property, the choice of the actual values given to these weights is subjective, just as is the choice of the underlying cause in the single-cause model. Hence, in the same way as the underlying cause is selected today using rules obtained by international consensus, in the future the choice of weights could be based on predetermined, possibly expert-based,[9] consensual strategies. For now, the weight-attribution issue needs to be addressed as a sensitivity analysis, as discussed next.

Weight-attribution Strategies

For a given death, let Π denote the weight attributed to the disease of interest. To determine Π, a first strategy is to distribute equal weights to each disease on the certificate, so that $\Pi = 1/(k+1)$ if the disease is mentioned on the certificate along with $k \geq 0$ other diseases, and $\Pi = 0$ if it is not mentioned (the "equal weights" approach). However, this strategy does not address requirement (3). In particular, similar to the any-mention and single-cause approaches, it ignores some expert, individual-level information contained in the certificate that can be used to infer the importance of the disease of interest relative to others. Specifically, four pieces of information are available: whether the disease is mentioned, whether it was selected as the underlying cause, the number of other diseases mentioned, and the position of the disease on the certificate.

A strategy that uses all this information and addresses requirement (3) is to attribute, whenever other diseases are present, a larger weight ω to the underlying cause, and distribute the rest of the weight equally among the other diseases mentioned. Conveniently, taking $\omega = 1$ coincides with the single-cause model. A refined version of this strategy, which we adopt in our simulation study and illustrative example, excludes diseases that are complications or consequences of the underlying cause, theoretically indicated in part I of the certificate, before assigning the weights; thus these diseases are assigned a zero weight. The rationale for this choice is that, when considering that death may be caused by exposure to several diseases simultaneously, it is more meaningful from a causal perspective to examine separate disease processes (i.e., diseases on separate causal pathways). Indeed, those diseases assigned a zero weight can be thought of as mediating factors on the path between the underlying cause and death.

Specifically, for a chosen ω, the strategy is as follows:

Rule 1: $\Pi = 1$ if the disease is the underlying cause and no other diseases are mentioned in part II of the certificate.

Rule 2: $\Pi = \omega$ if the disease is the underlying cause and other diseases are mentioned in part II of the certificate.

Rule 3: $\Pi = (1-\omega)/(k_2+1)$ if the disease is not the underlying cause but is mentioned in part II of the certificate along with $k_2 \geq 0$ other diseases (different from the underlying cause).

Rule 4: $\Pi = 0$ if the disease is not mentioned anywhere on the certificate or is mentioned in part I but is not the underlying cause.

In rule 3, the condition in parentheses is added because the underlying cause (as determined by WHO rules) may have been mentioned in part II of the certificate but should not be considered a "secondary" cause.

The value of ω being subjective, different values for ω need to be assessed. We recommend assessing the following four possibilities, or close variations of these: (1) $\omega = 1$ ("single-cause" approach), (2) $\omega = 0.75$, (3) $\omega = 0.5$, and (4) "equal weights" approach.

Multistate Model for Studying Disease-related Mortality

Once weights are attributed, the study of mortality related to a disease of interest can be based on the multistate model in Figure 2. All individuals begin at the "alive" state and, at death, move on to exactly one of several absorbing states, with state π $(0 \leq \pi \leq 1)$ representing death for which a weight π was attributed to the disease under consideration (i.e., $\Pi = \pi$). Although in theory the number of possible states (i.e., values of Π) is infinite, in practice we will deal with only a finite number, possibly small depending on the weight-attribution strategy.

This model extends the usual competing risks model (Figure 1) in that now there are competing endpoints

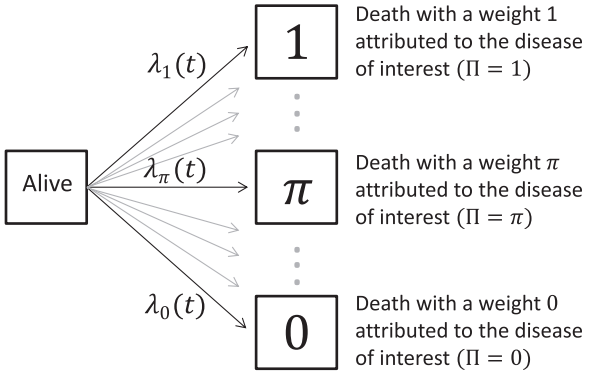

FIGURE 2. Multistate model for studying mortality related to the disease of interest based on the multiple-cause approach. The model has a potentially infinite number of absorbing states, one for each π such that $0 \leq \pi \leq 1$. The infinity of states is represented schematically in gray.

corresponding to death from a combination of diseases (states with $0 < \pi < 1$). Nonetheless, the statistical structure is equivalent, with the outcome for each individual being bivariate, now consisting of (T, Π). Hence, for each π $(0 \leq \pi \leq 1)$, it would be possible to use the same methods to study the state-specific hazard $\lambda_\pi(t)$, which is the instantaneous rate of transition into state π. While this possibility of studying end-points representing disease mixtures has been suggested before,[11,12] it has not found appeal as this will be relevant only in some specific cases (see "Discussion"). Designing public health policies that target death from disease combinations in general would be impractical and often senseless as there may be very few deaths for each given combination. In epidemiologic research, it seems unappealing to study exposure effects on the risk of death from a combination of diseases, specifically when each of these is considered to drive separate processes, arising as the individual ages. Rather, interest lies in goals 1 and 2, which can be addressed in this new framework as detailed next.

Addressing Goal 1

The total burden of mortality attributed to the disease of interest can be quantified in absolute terms using the sum of fractions of deaths attributed to that disease:

$$\frac{\sum_{\pi>0} \pi N_\pi}{\text{Total person-time at risk}}, \quad (1)$$

where N_π is the number of deaths for which $\Pi = \pi$. In another paper[6] we use (1) which is the maximum likelihood estimator of $\sum_{\pi>0} \pi \lambda_\pi$ if all hazards λ_π are assumed constant,[12] as a basis for estimating age- and sex-standardized disease-related mortality rates in France.

Addressing Goal 2

In Figure 2, the rates of transition into the "pure" states 1 and 0, $\lambda_1(t)$ and $\lambda_0(t)$, represent the forces of mortality

due solely to the disease of interest and other diseases, respectively. We call these functionals the *pure hazards*. For goal 2, it is necessary to identify a functional interpretable as the force of mortality driven by the disease process of interest, whether other separate disease processes are simultaneously present or not. We can endow $\lambda_1(t)$ with this broader interpretation by parameterizing the transition hazards in the multistate model in such a way that each state-specific hazard rate $\lambda_\pi(t)$ is completely determined by a combination of $\lambda_1(t)$ and $\lambda_0(t)$. Addressing goal 2 then amounts to regression analysis of $\lambda_1(t)$.

Specifically, letting $\mathbf{X}$ and $\mathbf{Z}$ denote (possibly overlapping) vectors of covariates (e.g., exposure status, gender, etc.) influencing $\lambda_1(t)$ and $\lambda_0(t)$, respectively, a possible assumption is that for each π $(0 \le \pi \le 1)$:

$$\lambda_\pi(t \mid \mathbf{X},\mathbf{Z}) = \pi\lambda_1(t \mid \mathbf{X}) + (1-\pi)\lambda_0(t \mid \mathbf{Z}). \quad \text{(A0)}$$

Assumption (A0) has the interpretation that the force driving individuals into state π results from a weighted sum of the forces of the two separate processes at play. Technically, fitting regression models for the pure hazards under (A0) ensures that all deaths partially attributed to the disease of interest ($\Pi > 0$) contribute information to the estimation of the effect parameters determining $\lambda_1(t)$, and the information contribution is (loosely speaking) proportional to the role the disease of interest played in those deaths. Other assumptions expressing $\lambda_\pi(t)$ in terms of $\lambda_1(t)$ and $\lambda_0(t)$ could be adopted if deemed suitable. Assumption (A0) has an intuitive interpretation, appeals to the additivity of rates, and results in a similar property for the cumulative incidence functions (see eAppendix, Section 1, http://links.lww.com/EDE/B77).

COX REGRESSION FOR THE PURE HAZARDS

Required Data

The required data for each individual $i = 1,\ldots,n$ are: the minimum of the time-to-death and a right-censoring time, $\tilde{T}_i = \min\{T_i, C_i\}$; the censoring indicator, $U_i = 1(C_i \le T_i)$; if the individual is uncensored ($U_i = 0$), the proportion Π_i of his death attributed to the disease of interest; and $\mathbf{X}_i$ and $\mathbf{Z}_i$, the covariates influencing the pure hazards of the disease of interest and other diseases, respectively (these could be time dependent, but we do not explicitly allow for this below). For identifiability it is necessary to observe some individuals in the pure states (i.e., with $\Pi_i = 1$ or 0).

Model

We assume independent and identically distributed data and independent right-censoring given covariates, (A0) and that both pure hazards follow Cox models and are related, as follows:

$$\lambda_1(t \mid \mathbf{X}) = \lambda_{10}(t)\exp(\boldsymbol{\rho}'\mathbf{X}) \quad \text{(A1)}$$

$$\lambda_0(t \mid \mathbf{Z}) = \lambda_{00}(t)\exp(\boldsymbol{\phi}'\mathbf{Z}) \quad \text{(A2)}$$

$$\lambda_{00}(t) = \lambda_{10}(t)\exp\{-\xi(t)\} \quad \text{(A3)}$$

The apostrophe denotes transposition. Under (A0)–(A3), it follows that

$$\lambda_\pi(t \mid \mathbf{X},\mathbf{Z}) = \lambda_{10}(t)[\pi\exp\{\boldsymbol{\rho}'\mathbf{X}\} + (1-\pi)\exp\{-\xi(t) + \boldsymbol{\phi}'\mathbf{Z}\}].$$

Generally the target parameter is $\boldsymbol{\rho}$, the effect of $\mathbf{X}$ on the pure hazard of the disease of interest. Yet, unlike in the single-cause model, we need to posit a model for $\lambda_0(t \mid \mathbf{Z}_i)$ as well, for example as in (A2), and we need to fit both models simultaneously to incorporate deaths going into the mixture states. This is the consequence of (A0), which involves both pure hazards for deaths with $0 < \Pi_i < 1$. As it stands, (A3) does not impose any restrictions. However, the specific estimation theory used here requires a fully parametric model for the log ratio of the baseline pure hazards, $\xi(t)$. In our example, we parameterized this as a piecewise constant function. Such models provide much flexibility while keeping estimating procedures simpler.

We fitted this model using an estimating equation approach,[14,15] inspired from the literature on missing and misclassified causes of death[16–19] (details in eAppendix, Section 2, http://links.lww.com/EDE/B77). We derived a quantity reflecting the absolute burden on mortality of the disease of interest in the baseline population that can be estimated from the fitted model, and can be seen as a cumulative/time-dependent version of (1). We call it the disease-attributed cumulative baseline hazard (see eAppendix, Section 3, http://links.lww.com/EDE/B77).

SIMULATION STUDY

We considered the simplified scenario where the disease of interest could be: not mentioned on the certificate, selected as underlying cause or mentioned in part II along with up to two diseases, and assumed weight-attribution strategy (2) ($\omega = 0.75$). This results in a multistate model with six absorbing states (see eAppendix, Section 4, http://links.lww.com/EDE/B77), from which data can be generated by completely specifying all the transition hazards.[20] We assumed (A0)–(A3) and that the log ratio of the pure baseline hazards was constant ($\xi(t) = \xi$). Specifically, for each π, $\lambda_\pi(t \mid X) = 0.002t\{\pi\exp(\rho X) + (1-\pi)\exp(-\xi + \phi X)\}$, where X was a binary exposure. We superimposed around 30% independent right censoring. Simulation R code is provided in the eAppendix (Section 7, http://links.lww.com/EDE/B77).

We compared various approaches to estimate the target parameter ρ: Cox regression for the pure hazards based on correct ($\omega = 0.75$) and misspecified ($\omega = 0.5$, "equal weights") weight-attribution strategies, and the single-cause ($\omega = 1$) and any-mention approaches, both implemented using standard Cox regression considering, respectively, deaths with the disease as underlying cause and any mention of the disease ($\Pi > 0$) as events of interest.

Figure 3 shows the distribution of the bias $\hat{\rho} - \rho$ across different scenarios (see eAppendix, Section 5, for other scenarios, http://links.lww.com/EDE/B77). The multiple-cause model with the correct weight-attribution strategy ($\omega = 0.75$) was unbiased, as expected, and only moderately biased with misspecified strategies ($\omega = 0.5$ and "equal weights"). This approach was the least precise as the power here is driven by a reduced number of effective events, those considered "pure", and it involves estimation of one more parameter (ξ). Still, overall this approach provided a gain in terms of mean squared error in all scenarios considered, even under misspecified weight-attribution strategies (see eAppendix, Section 5, http://links.lww.com/EDE/B77).

The single-cause and any-mention approaches yielded unbiased results when $\rho = \phi$, but bias arose as these effects diverged. The bias in these two approaches is due to the effects on each of the pure hazards counterbalancing due to the consideration of deaths with $\Pi = 0.75$ and $0 < \Pi < 1$, respectively, as entirely due to the disease of interest, highlighting the importance of requirement (3). The any-mention approach resulted as foreshadowed in highly precise estimates.

EXAMPLE: EDUCATIONAL INEQUALITIES IN MORTALITY

The goal was to quantify relative educational inequalities in disease-related mortality based on the cohort of men alive and ages ≥30 years on 01 January 2000 (n = 148,384) in a cross-sectionally representative 1% sample of the French population, which was linked with the French National Cause-of-Death Register.[20,21] After an 8-year follow-up, at the administrative censoring date (December 31, 2007), 11.8% had died (n = 17,512). This study was approved by the French data protection committee and institutional ethical review board (Commission Nationale de l'Informatique et des Libertés, authorisation No. 902368v2).

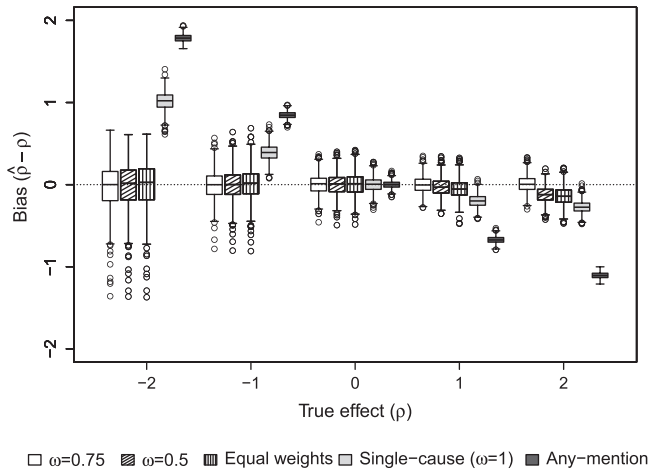

FIGURE 3. Simulation study results: box plots of the bias distribution for five approaches across 1,000 datasets of $n = 4{,}000$ individuals, generated assuming $\xi = -1$, $\phi = 0$ and varying values of ρ.

For each disease group separately, we fitted Cox models for mortality due to the disease group of interest and other diseases considering the four recommended weight-attribution strategies and the any-mention approach. The parameter of interest was the relative index of inequality, which is the exponentiated regression coefficient of the socioeconomic rank, a variable ranging from 0 = most educated to 1 = least educated.[22] Age was the time scale in all approaches, and for strategies (2)–(4), $\xi(t)$ was modeled as piecewise constant, with pieces corresponding to 10-year age-groups.

Table 1 shows the distribution of event types. The ratio AM/UC (=deaths with disease mentioned on certificate/deaths with disease selected as underlying cause) is used in the literature to identify diseases for which multiple-cause approaches are more relevant.[23] Table 2 shows estimates of the inequality index associated with the pure hazard of the disease group of interest (RII_1). Estimates obtained for the other pure hazard were similar across all disease groups and approaches, with $\widehat{\mathrm{RII}}_0 \approx 2$, and $\xi(t)$ was generally negative across the age scale (see eAppendix, Section 6, http://links.lww.com/EDE/B77).

More sensitivity to the method of estimation was observed for diseases with higher AM/UC ratio and an effect $\rho = \log(\mathrm{RII}_1)$ more discrepant from $\phi = \log(\mathrm{RII}_0) \approx \log(2)$, as in the simulations. For diseases with $\rho \approx \phi$ (endocrine and nutritional, musculoskeletal, cardiovascular, neoplasms, and other diseases) all approaches yielded similar estimates of RII_1. For diseases of the nervous system and the sense organs, for which $\rho < \phi$, the any-mention approach yielded a slightly higher estimate. For all other diseases $\rho > \phi$, and the single-cause and any-mention approaches yielded lower estimates. The multiple-cause model generally yielded similar results across the three remaining weight-attribution strategies, and for some disease-groups revealed larger socioeconomic inequalities, a result that may be of public heath relevance. As expected, these approaches yielded wider confidence intervals, particularly for diseases with a small number of events (blood and skin) for which results were more unstable across approaches.

The ranking of disease groups by their burden on mortality as determined by disease-attributed and cause-specific cumulative baseline hazard curves derived from these models was found to be sensitive to the approach used in some age-groups (see eAppendix, Section 6, http://links.lww.com/EDE/B77). Different approaches may thus lead to establishing different public health priorities.

DISCUSSION

Our proposal enables the use of multiple cause of death data to better understand the etiology and burden of disease-related mortality. This valuable source of data coupled with the proposed methods could provide new insights, especially for diseases with high AM/UC ratios.

TABLE 1. Number of Deaths for Each Event Type by Disease Group Among the Deceased Individuals (n = 17,512)

Disease Group[a]	ICD-10 Codes	UC	AM	AM/UC	Detailed Event Types					
					Ev1	Ev2	Ev3	Ev4	Ev5	Ev6
Neoplasms	C00-D48	6,227	6,911	1.1	4,931	1,296	269	139	123	10,754
Cardiovascular	I00-I99	4,671	8,120	1.7	3,165	1,506	591	433	310	11,507
Digestive	K00-K93	840	1,930	2.3	532	308	89	61	68	16,454
Nervous/sense	G00-H95	650	1,519	2.3	463	187	174	128	130	16,430
Musculoskeletal	M00-M99	78	234	3.0	49	29	25	33	35	17,341
Respiratory	J00-J99	1,118	3,826	3.4	650	468	200	186	199	15,809
Endocrine/nutritional	E00-E90	531	1,884	3.5	343	188	233	291	329	16,128
Mental	F00-F99	484	1,847	3.8	328	156	403	231	201	16,193
Infectious	A00-B99	307	1,330	4.3	160	147	33	43	40	17,089
Genitourinary	N00-N99	254	1,297	5.1	148	106	94	147	159	16,858
Skin	L00-L99	38	265	7.0	22	16	15	31	41	17,387
Blood	D50-D89	57	433	7.6	33	24	32	33	35	17,355
Other	Other	2,257	10,314	4.6	1,943	314	492	395	393	13,975

[a]Disease groups sorted by increasing AM/UC ratio, except for group "other."

AM, any mention of the disease on the certificate; Ev1, disease is underlying cause and no other diseases mentioned in part II; Ev2, disease is underlying cause and other diseases mentioned in part II; Ev3, disease is not underlying cause but is mentioned in part II by itself; Ev4, disease is not underlying cause but is mentioned in part II with one other disease; Ev5, disease is not underlying cause but is mentioned in part II with two or more diseases; Ev6, disease is not mentioned or is mentioned in part I but is not the underlying cause; ICD-10, International Classification of Diseases and Related Health Problems, 10th revision; UC, Disease selected as underlying cause.

TABLE 2. For Each Disease Group, Estimates (and 95% Confidence Intervals) of the Relative Index of Inequality in Disease-related Mortality ($R\hat{I}I_1$) Obtained Using Various Approaches

Disease Group[a]	Any Mention	Single Cause	Multiple Cause		
		$\omega = 1$	$\omega = 0.75$	$\omega = 0.5$	Equal Weights
Neoplasms	1.8 (1.6, 1.9)	1.8 (1.6, 2.0)	1.7 (1.6, 1.9)	1.7 (1.6, 1.9)	1.7 (1.5, 1.9)
Cardiovascular	2.0 (1.9, 2.2)	2.1 (1.9, 2.3)	2.0 (1.8, 2.3)	2.0 (1.8, 2.3)	2.0 (1.8, 2.3)
Digestive	2.7 (2.2, 3.2)	3.3 (2.5, 4.4)	4.7 (3.3, 6.7)	4.9 (3.4, 7.0)	4.9 (3.4, 7.0)
Nervous/sense	1.3 (1.1, 1.6)	1.0 (0.7, 1.3)	1.0 (0.7, 1.4)	1.0 (0.7, 1.4)	1.0 (0.7, 1.5)
Musculoskeletal	2.1 (1.3, 3.6)	2.0 (0.8, 4.9)	2.0 (0.6, 6.2)	2.1 (0.7, 6.5)	1.8 (0.6, 5.7)
Respiratory	2.4 (2.1, 2.7)	3.1 (2.4, 4.0)	3.5 (2.5, 4.8)	3.6 (2.6, 5.0)	3.5 (2.5, 5.0)
Endocrine/nutritional	2.2 (1.8, 2.6)	2.4 (1.7, 3.4)	2.3 (1.5, 3.5)	2.2 (1.5, 3.4)	2.2 (1.5, 3.5)
Mental	3.5 (2.9, 4.2)	2.9 (2.0, 4.3)	3.5 (2.2, 5.4)	3.9 (2.5, 6.0)	3.9 (2.5, 6.1)
Infectious	2.0 (1.6, 2.5)	1.9 (1.2, 3.0)	2.6 (1.4, 5.0)	2.7 (1.4, 5.0)	2.7 (1.4, 5.0)
Genitourinary	1.9 (1.5, 2.4)	2.5 (1.5, 4.2)	2.8 (1.4, 5.7)	2.8 (1.4, 5.7)	2.8 (1.4, 5.7)
Skin	2.8 (1.7, 4.7)	5.8 (1.2, 27.4)	5.8 (0.7, 44.9)	3.5 (0.5, 23.9)	4.4 (0.6, 31.9)
Blood	2.1 (1.4, 3.1)	3.8 (1.2, 12.1)	8.9 (1.6, 48.5)	8.9 (1.6, 48.3)	9.0 (1.7, 48.9)
Other	2.0 (1.9, 2.2)	1.9 (1.6, 2.2)	1.9 (1.6, 2.2)	1.9 (1.6, 2.2)	1.8 (1.6, 2.2)

[a]Disease groups sorted as in Table 1.

As mentioned, the notion of a set of weights with the described properties is only an approximation of the way diseases cause mortality. The problem of weight attribution can be paralleled to the classic epidemiologic problem of attributing causal responsibility to exposures when considering multiple exposures that may interact in the "sufficient cause" sense.[24] For example, in the extreme scenario, where two diseases can cause death only when present together but not on their own, death can be considered to be 100% attributable to each of these in that the death would be prevented by removing any of the two diseases.[25,26] If there is such a strong interaction between two or more diseases, then it seems reasonable to define this configuration of conditions as a "disease" on its own, and actually the ICD is moving toward having specific "combination" codes for such cases.[27] Also, arguably the study of disease-related mortality is restricted to "diseases" than can actually cause death, and accordingly the proposed approach requires that the disease

of interest is sufficient in some cases to cause death through the requirement of "pure events."

Outside this setting of strong interactions, when considering diseases that are sometimes, but not always, sufficient to cause death, interest will generally be on studying each disease separately as discussed previously. This requires the distribution of causal responsibility among the diseases on the certificate, and the three approaches compared here suffer from the subjectivity associated with this enterprise. The any-mention approach attributes a weight of 100% to each of these, which does not make sense from an epidemiologic point of view in this scenario, when considering that some diseases on the certificate may have been sufficient but were maybe not necessary to cause death. That is, the death may or may not be prevented (or delayed) by removing the disease. The single-cause approach attributes 100% to the underlying cause and 0% to all other diseases, ignoring them. The method proposed here is based on a set of weights adding up to one, which is also subjective, but performing a sensitivity analysis as described provides a structured way to approach this subjectivity. Importantly, the proposed methods can be used even if weights adding up to more than one are attributed, although allowing for this possibility further complicates the weight-attribution issue. Also, the sum of the disease-attributed cumulative hazards will then add up to more than the total cumulative hazard of mortality, which complicates interpretation but this could be considered a natural requirement of "shared causal responsibility".[25]

In the future, the choice of weights may be the object of an international consensus as is the choice of the underlying cause today. One could also consider estimating these weights as population attributable fractions[28] or related quantities,[29] but these also suffer from subjectivity in the way causal responsibility is distributed among the different diseases.[30] Furthermore, this would require data from a large longitudinal cohort, recording the incidence of all diseases and mortality over time, which is generally unavailable. Another unappealing aspect of replacing the weights by population estimates is the loss of the individual-level death certificate information, particularly given the arguably strong causal nature of diseases mentioned at the moment of death.

The acknowledgment of multiple-cause mortality does not solve the "competing risks" problem in the causal sense. For each individual, we still observe only the time-to-death from one combination of diseases, and their potential time-to-death from another combination is not observable. Thus, the independence between these potential times-to-death remains unidentifiable, and so does the actual target quantity for goal 2: the causal "marginal" hazard (i.e., the hazard of deaths caused by the disease of interest in a counterfactual world without other diseases).[30] As noted by Prentice and colleagues,[9] what multiple cause data enable is the identifiability of interrelations between diseases in causing death because we observe individuals dying from combinations of these. Importantly, neither the approaches proposed here nor those described for the single-cause model, make any assumptions regarding the (in)dependence between the potential times-to-death. Arguably, the pure hazard, being the rate of deaths caused exclusively by the disease of interest, is a quantity that is conceptually closer to the marginal hazard than the cause-specific hazard. Still, whether the pure and marginal hazards coincide continues to depend on this unidentifiable independence assumption.

We mentioned that it is more meaningful from a causal perspective to consider diseases on separate causal pathways. The current death certificate allows us to identify these only to a limited extent due to its design. In addition, the well-known issues of quality and comparability of death certificate data across countries and over time become even more important when one starts to consider all the diseases and their position.[4,5] One step to improve the accuracy of these data would be to devise a set of international rules, in the spirit of those for selecting the underlying cause, with the aim of identifying separate morbid processes in the death certificate and selecting their initiating causes.

The findings in the example rely on assumptions (A0)–(A3) holding, which points to the need for methods to assess the fit of this model to the data, as well as models that consider other such assumptions. Overall, further research into the statistical modeling of multiple cause of death data is warranted; despite their many drawbacks, mortality databases remain a reference for public health monitoring and life-course epidemiology as they currently represent the only data that is systematically collected for every individual in most countries by an expert and in a standardized way.

ACKNOWLEDGMENTS

The authors thank John B. Carlin and the reviewers for their helpful comments.

REFERENCES

1. World Health Organisation. *International Statistical Classification of Diseases and Related Health Problems*. 10th Revision. Volume 2. Second Edition. Geneva: World Health Organization; 2004. Available at: http://www.who.int/classifications/icd/ICD-10_2nd_ed_volume2.pdf. Accessed June 16, 2016.
2. World Health Organization. *International Statistical Classification of Diseases and Related Health Problems*. 10th Revision. Volume 2. *Instruction manual*. Geneva: World Health Organization; 2010.
3. Omram AR. The epidemiologic transition: a theory of the epidemiology of population change. *Milbank Quarterly*. 2005:83:731–757.
4. Israel RA, Rosenberg HM, Curtin LR. Analytical potential for multiple cause-of-death data. *Am J Epidemiol*. 1986;124:161–179.
5. Désesquelles AF, Salvatore MA, Pappagallo M, et al. Analysing multiple causes of death: which methods for which data? An application to the cancer-related mortality in France and Italy. *Eur J Population*. 2012;28:467–498.
6. Piffaretti C, Moreno-Betancur M, Lamarche-Vadel A, Rey G. A method for measuring the level of cause-related mortality using weighted multiple cause of death data. *Bulletin of the World Health Organization. Forthcoming*.
7. Goodman RA, Manton KG, Nolan TF Jr, Bregman DJ, Hinman AR. Mortality data analysis using a multiple-cause approach. *JAMA*. 1982;247:793–796.

8. Redelings MD, Sorvillo F, Simon P. A comparison of underlying cause and multiple causes of death: US vital statistics, 2000–2001. *Epidemiology*. 2006;17:100–103.
9. Prentice RL, Kalbfleisch JD, Peterson AV Jr, Flournoy N, Farewell VT, Breslow NE. The analysis of failure times in the presence of competing risks. *Biometrics*. 1978;34:541–554.
10. Bakal JA, Westerhout CM, Armstrong PW. Impact of weighted composite compared to traditional composite endpoints for the design of randomized controlled trials. *Stat Methods Med Res*. 2015;24:980–988.
11. Manton KG, Tolley DH, Poss SS. Life table techniques for multiple-cause mortality. *Demography*. 1976;13:541–564.
12. Chatterjee N, Sinha S, Diver WR, Feigelson HS. Analysis of cohort studies with multivariate and partially observed disease classification data. *Biometrika*. 2010;97:683–698.
13. Grambauer N, Schumacher M, Dettenkofer M, Beyersmann J. Incidence densities in a competing events analysis. *Am J Epidemiol*. 2010;172:1077–1084.
14. Goetghebeur E, Ryan L. Analysis of competing risks survival data when some failure types are missing. *Biometrika*. 1995;82:821–833.
15. Van Rompaye B, Jaffar S, Goetghebeur E. Estimation with Cox models: cause-specific survival analysis with misclassified cause of failure. *Epidemiology*. 2012;23:194–202.
16. Ebrahimi N. The effects of misclassification of the actual cause of death in competing risks analysis. *Stat Med*. 1996;15:1557–1566.
17. Lu KF, Tsiatis AA. Multiple imputation methods for estimating regression coefficients in the competing risks model with missing cause of failure. *Biometrics*. 2001;57:1191–1197.
18. Moreno-Betancur M, Latouche A. Regression modeling of the cumulative incidence function with missing causes of failure using pseudovalues. *Stat Med*. 2013;32:3206–3223.
19. Moreno-Betancur M, Rey G, Latouche A. Direct likelihood inference and sensitivity analysis for competing risks regression with missing causes of failure. *Biometrics*. 2015;71:498–507.
20. Beyersmann J, Latouche A, Buchholz A, Schumacher M. Simulating competing risks data in survival analysis. *Stat Med*. 2009;28:956–971.
21. Couet C. INSEE's permanent demographic sample (EDP). *Courrier des statistiques, English series*. 2007;13:29–38.
22. Moreno-Betancur M, Latouche A, Menvielle G, Kunst AE, Rey G. Relative index of inequality and slope index of inequality: a structured regression framework for estimation. *Epidemiology*. 2015;26:518–527.
23. Steenland K, Nowlin S, Ryan B, Adams S. Use of multiple-cause mortality data in epidemiologic analyses: US rate and proportion files developed by the National Institute for Occupational Safety and Health and the National Cancer Institute. *Am J Epidemiol*. 1992;136:855–862.
24. Rothman KJ, Greenland S, Lash TL. *Modern Epidemiology*. Philadelphia: Lippincott Williams & Wilkins; 2008.
25. Poole C. A history of the population attributable fraction and related measures. *Ann Epidemiol*. 2015;25:147–154.
26. Hernan MA, Robins JM. *Causal Inference.* Boca Raton, FL: Chapman and Hall/CRC; 2016.
27. World Health Organization. *International Classification of Diseases*. 10th Revision, Clinical Modification (ICD-10-CM)-Official Guidelines for Coding and Reporting FY-2016. Available at: http://www.cdc.gov/nchs/data/icd/10cmguidelines2016Final.pdf. 2016. Accessed December 15, 2015.
28. Benichou J. A review of adjusted estimators of attributable risk. *Stat Methods Med Res*. 2001;10:195–216.
29. Land M, Vogel C, Gefeller O. Partitioning methods for multifactorial risk attribution. *Stat Methods Med Res*. 2001;10:217–230.
30. Tsiatis AA. A nonidentifiability aspect of the problem of competing risks. *Proc Natl Acad Sci.* 1975;72:20–22.