Cause-of-death dependencies: structure, evolution by age and sensitivity to disruptions

Elizabet Ukolova and Trifon I. Missov

Introduction

Typically, mortality analysis by causes are conducted for only 35% of the death records listed on death certificates (Authors calculations, data source: NCHS, 2023). This is primarily because research predominantly emphasizes the single underlying cause of death, despite the fact that nearly 80% of deaths in the USA result from more than one cause. Focus on single underlying causes of death has led to a restricted understanding of the interconnections between causes of death. We are left with essential questions: What is the structure of cause of death dependencies? How does this structure evolve with age? How would it change if hypothetical disruptions in flows of deaths between diseases would be embedded into the structure of cause of death dependencies?

While relationships among causes of death have been thoroughly studied [1–6] the age dimension has often been overlooked. Furthermore, to date, there have been no studies addressing how the dissolution of selected significant associations can affect the structure of associations among causes of death. The current contribution has three primary objectives: (i) to identify the most frequent cause-of-death pairs and to measure the strength of associations among them, (ii) to study these dependencies by age at death, and (iii) to explore how the structure of dependencies among diseases would change if the most frequent cause-of-death pairs were indeed independent.

Data

We use the publicly available Mortality multiple cause data files for the US in 2019. We selected this year for our analysis to avoid working with cause of death data during the COVID-19 pandemic, as cause of death coding during this period may be biased. The Multiple Cause of Death Mortality Data was obtained from the National Bureau of Economic Research [7], which compiles microdata on mortality from the National Vital Statistics System of the National Center for Health Statistics [8]. We analyze the dependencies among the 16 major causes of death grouped in accordance with the Cause of death shortlist proposed by the Human Cause of Death Database [9].

Methods

First, we quantify the strength of associations among causes of death using odds ratios (OR), the cause of death association indicator (CDAI) and conditional probabilities of having disease *j* given a person has disease *i*. Subsequently, we explore how these associations change with age. We then employ both OR and CDAI as similarity matrices to identify disease clusters using hierarchical clustering and principal component analysis. In the final step, we introduce perturbations into the

structure of causes of death by optimizing the marginal distributions in contingency tables to make the chi-squared test result indicate "independence", as illustrated in Table 1. Clearly, the optimization of marginal distributions in the contingency table for diseases *i* and *j* has the potential to influence all other associations involving either disease *i* or disease *j*.

Table 1: Scheme of implementation of distortions into cause of death dependencies

Disease i: Heart disease, disease j: Endocrine, nutritional and metabolic diseases

		Disease j YES	Disease j NO	
Disease i	YES	376 200	1 052 256	1 428 456
	NO	123 205	1 309 862	1 433 067
		499 405	2 362 118	2 861 523

	Observed	Adjusted
Odds ratio	3.8010	1.0063

Former X^2 test:

	$D_{\neg i\neg j}$	$D_{\neg ij}$	$D_{i\neg j}$	D_{ij}	To be optimized $D_{\neg i}$	D_i	$D_{\neg j}$	D_j	Constrain Total	Total
Observed	1 309 862	123 205	1 052 256	376 200	**1 433 067**	**1 428 456**	**2 362 118**	**499 405**	2 861 523	2 861 523
Expected	1 182 962	250 105	1 179 156	249 300	1 433 067	1 428 456	1 433 067	1 428 456	2 861 523	2 861 523
X2	13 613	64 387	13 657	64 595						
SumX2	**156 252**	**Target value**								
P-value	<0.000									

X^2 test adjusted for independence:

	$D_{\neg i\neg j}$	$D_{\neg ij}$	$D_{i\neg j}$	D_{ij}	To be optimized D¬i	Di	D¬j	Dj	Constrain Total	Total
Observed	936 213	279 305	1 269 805	376 200	**1 215 385**	**1 646 005**	**2 206 018**	**655 505**	2 861 523	2 861 523
Expected	936 970	278 415	1 268 945	377 060	1 215 385	1 646 005	1 215 385	1 646 005	2 861 390	2 861 390
X2	0.612	2.845	0.582	1.960						
SumX2	**6**	**Target value**								
P-value	0.1117									

Source: Author

Preliminary findings

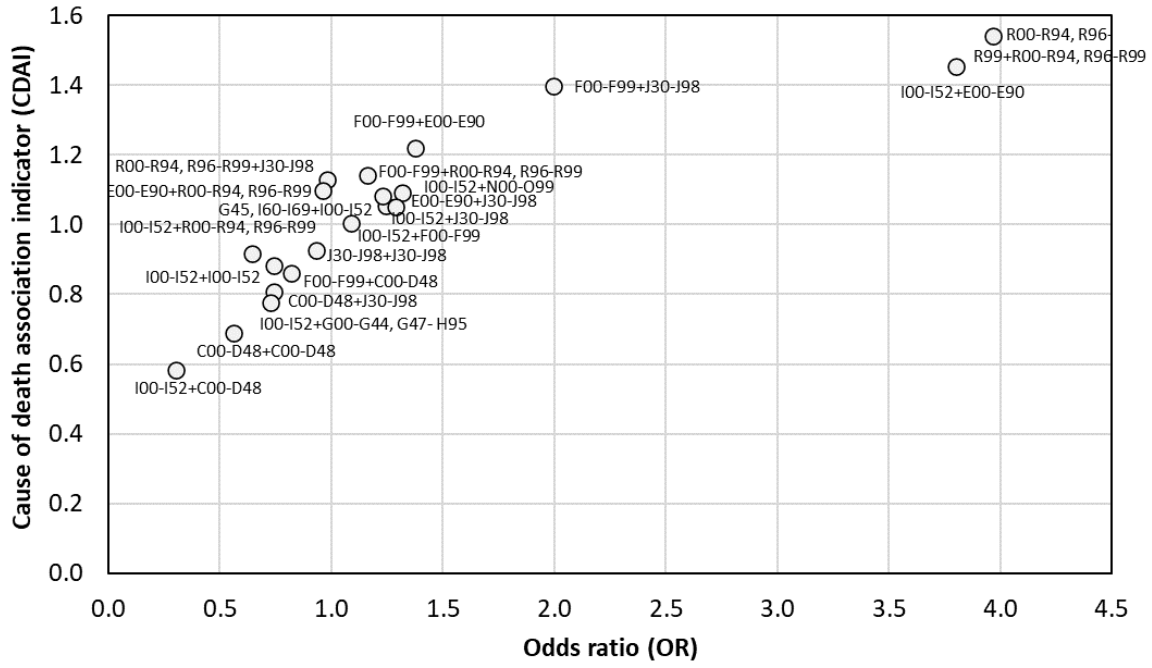
In 2019, nearly 65% of deaths in the US were recorded with at least one of the following cause of death pairs presented in the Table 2. Most of these pairs were significantly associated with each other, which show both the OR and CDAI, that are strongly correlated with each other, as shown in Figure 1. However, CDAI is more concentrated around unity, which indicates independence between causes of death, whereas OR exaggerates the strengths of associations. The disparity is especially pronounced in symptoms and signs (R00–R99, excluding R95) with heart diseases (I00–I52) and in endocrine, nutritional and metabolic diseases (E00–E90) with heart diseases (I00–I52). The associations show a specific age profile, as plotted in Figure 2 with conditional probabilities. Starting with age ten, the probability of having symptoms and signs decreases with age, which is especially pronounced in pairs involving mental and behavioral diseases and heart diseases. In contrary, conditional probabilities of having heart disease and cerebrovascular diseases increase with age.

Table 2: 20 leading cause of death pairs, USA, 2019, both sexes combined

	Cause 1	Cause 1 codes	Cause 2	Cause 2 codes
1	Heart diseases	I00–I52	Heart diseases	I00–I52
2	Heart diseases	I00–I52	Endocrine, nutritional and metabolic diseases	E00–E90
3	Symptoms and signs	R00–R94, R96–R99	Symptoms and signs	R00–R94, R96–R99
4	Heart diseases	I00–I52	Other respiratory diseases	J30–J98
5	Heart diseases	I00–I52	Mental and behavioral disorders	F00–F99
6	Heart diseases	I00–I52	Symptoms and signs	R00–R94, R96–R99
7	Mental and behavioral disorders	F00–F99	Other respiratory diseases	J30–J98
8	Heart diseases	I00–I52	Neoplasms	C00–D48
9	Mental and behavioral disorders	F00–F99	Symptoms and signs	R00–R94, R96–R99
10	Symptoms and signs	R00–R94, R96–R99	Other respiratory diseases	J30–J98
11	Heart diseases	I00–I52	Diseases of the genitourinary system	N00–O99
12	Heart diseases	I00–I52	Dis. of the nervous sys. and the sense organs	G00–G44, G47– H95
13	Other respiratory diseases	J30–J98	Other respiratory diseases	J30–J98
14	Cerebrovascular diseases	G45, I60–I69	Heart diseases	I00–I52
15	Mental and behavioral disorders	F00–F99	Neoplasms	C00–D48
16	Mental and behavioral disorders	F00–F99	Endocrine, nutritional and metabolic diseases	E00–E90
17	Endocrine, nutritional and metabolic diseases	E00–E90	Symptoms and signs	R00–R94, R96–R99
18	Endocrine, nutritional and metabolic diseases	E00–E90	Other respiratory diseases	J30–J98
19	Neoplasms	C00–D48	Other respiratory diseases	J30–J98
20	Neoplasms	C00–D48	Neoplasms	C00–D48

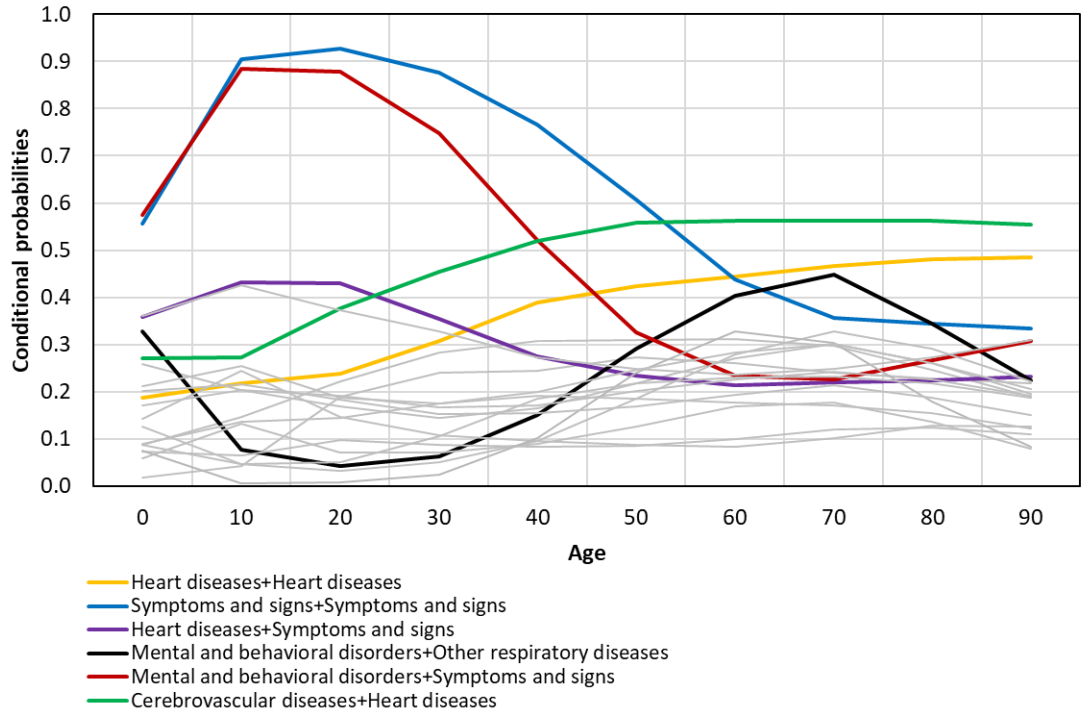
Authors calculations based on data source: NCHS (2023)

Figure 1: Relationship between odds ratio and cause of death association indicator



Authors calculations based on data source: NCHS (2023)

Figure 2: Age-specific conditional probabilities of having disease j given a person has disease i for leading cause of death pairs, USA, 2019, both sexes combined



Authors calculations based on data source: NCHS (2023)

Finally, we present expected results from the implementation of disturbances into associations. If we gradually optimize the results of the chi-squared test, so that it indicates the absence of significant associations (OR = 1) between two causes of death *i* and *j*, most pronounced changes can be found in associations with heart diseases, other respiratory diseases and mental and behavioral disorders. The magnitude of the change in odds ratio (OR) depends on the specific diseases under study. The most substantial impact is seen when dissolving associations between heart diseases and endocrine, nutritional, and metabolic disorders, resulting in an almost twofold increase in the risk of death for individuals with both heart disease and symptoms and signs, as well as those with heart disease and other respiratory conditions.

Conclusion

In 2019, deaths resulting from multiple causes were more common than single-cause deaths in the USA. Our preliminary findings identify age-specific patterns in cause-of-death relationships, and we contribute to the understanding of the interconnections between the causes of death by examining the effect that the breakdown of the associations between causes *i* and *j* can have on the association between causes *i* and *k*.

In a further step, we aim to check the robustness of our results by experimenting with various algorithms to find the optimal solution for the chi-squared test, along with testing different thresholds for statistical significance. In addition, we will expand the analysis to examine the role of age in disrupting dependencies among causes of death.

References

1. Egidi V, Salvatore MA, Rivellini G, D'Angelo S. A network approach to studying cause-of-death interrelations. Demogr Res. 2018;38:373–400.
2. Desesquelles A, Demuru E, Salvatore MA, et al.. Mortality from Alzheimer's disease, Parkinson's disease, and dementias in France and Italy: a comparison using the multiple cause-of-death approach. J Aging Health. 2014;26:283–315.
3. Mannino DM, Brown C, Giovino GA. Obstructive lung disease deaths in the United States from 1979 through 1993. An analysis using multiple-cause mortality data. Am J Respir Crit Care Med. 1997;156(3 Pt 1):8148148–8148818.
4. Bah S, Rahman MM. On measures of association for multiple cause mortality: do we need more measures? Canadian Stud Popul. 2011;38:93–104.
5. Rodriguez F, Blum MR, Falasinnu T, et al.. Diabetes-attributable mortality in the United States from 2003 to 2016 using a multiple-cause-of-death approach. Diabetes Res Clin Pract. 2019;148:169–178.
6. Moussa MA, El Sayed AM, Sugathan TN, Khogali MM, Verma D. Analysis of underlying and multiple-cause mortality data. Genus. 1992;48:89–105.
7. NBER, 2023. Multiple mortality data vital statistics. Available: https://www.nber.org/research/data/mortality-data-vital-statistics-nchs-multiple-cause-death-data
8. NCHS, 2023. Multiple mortality public use data. Available: https://www.cdc.gov/nchs/nvss/mortality_public_use_data.htm
9. Human Cause-of-Death Database. French Institute for Demographic Studies (France) and Max Planck Institute for Demographic Research (Germany). Cause of death Short list. Available at www.causeofdeath.org