**Assessing regional mortality disparities across intra- and international borders in Europe: *An application of the Earth Mover's Distance***

Laura Cilek[1], Markus Sauerberg[2], Pavel Grigoriev[1], Sebastian Klüsener[1]

1 The Federal Institute for Population Research (BiB; Wiesbaden, Germany)
2 Cancer Registry Hamburg (Süderstraße 30, Hamburg, 20097, Germany)

**Short Abstract**

Whereas the majority of studies on mortality trends and differentials tends to focus on countries and sub-national regions, less attention has been given to disparities between adjacent regions divided by intra- or international borders. We demonstrate the potential of such analyses through the application of the Earth Mover's Distance, a measure of statistical distance that quantifies the difference between two distributions with a single measure and can incorporate multiple dimensions within these distinct distributions. We show the one-dimensional application of the method, using mortality data for 526 European subregions to examine differences in the distribution of deaths by age across 179 international and 1073 domestic borders. We then empirically demonstrate its two-dimensional application, focusing on Austria, Czechia, and Slovakia using both age and cause of death. As expected, our results highlight the existence of large mortality differences along international borders, especially between countries separated by the former Iron Curtain. However, large mortality differences between adjacent regions can also occur within countries or along so-called cultural borders. Our innovative approach to quantitatively study cross-border disparities highlights new perspectives to study and understand spatial, socioeconomic, and other differences in the context of demographic research.

**Assessing regional mortality disparities across intra- and international borders in Europe: *An application of the Earth Mover's Distance***

Laura Cilek, Markus Sauerberg, Pavel Grigoriev, Sebastian Klüsener

The Federal Institute for Population Research (BiB; Wiesbaden, Germany)

Extended Abstract

**Background**

There is a growing interest towards examining sub-regional patterns and variations in mortality. Notably, within and between European countries, discernible east-west and north-south gradients have been identified and subjected to analysis over time (for example, Hrzic et al, 2023; Rau and Schmertmann, 2020 (Germany); Bramajo et al, 2023 (Spain); Bonnet and d'Albis, 2020 (France)). This more in-depth type of geographic consideration offers a more nuanced understanding of mortality differentials and equips researchers and policymakers with valuable insights into how regional distinctions, such as health policies, educational attainment, socioeconomic status, and other structural and environmental factors, may influence health and longevity within a region (see, for instance, Mühlichen et al 2023.). Nevertheless, these studies often focus around broader national or continental contexts, rather than seeking to pinpoint the evolution and causes of specific divergences between adjoining regions.

In this paper, we introduce the Earth Mover's Distance (EMD, also can be known as the Wasserstein metric, Kantorovich–Rubinstein metric, or Mallows's distance) as a new method to quantify dissimilarity between two mortality distributions. As a distance measure, the EMD is a solution to the optimal transport problem and can broadly be imagined as the amount of work needed to move one pile of dirt so that it exactly replicates another (dirt to be moved + distance moved). The measure can be calculated in a one or multidimensional context, meaning that we can include complex data within the compared distributions, such as causes of death, in our analysis. The result is a single number that quantifies the total difference between two distributions and their dimensions, meaning that it is digestible for policy makers and other non-demographers to understand. Our analysis seeks to enhance our comprehension of the differences in mortality between directly bordering regions.

**The Earth Mover's Distance as a measure of distributional differences**

Mathematically, the earth mover's distance is designed as a solution of the optimal transport problem proposed by Monge (1781), meaning that it calculates the minimum "work" needed to transform one distribution so that it replicates exactly a second distribution (e.g. Vallender 1974, Rubner et al, 2000). If thought of in the context of two piles containing the same, but differently distributed amount of dirt, the measure can be informally envisioned as the minimum amount of dirt that must be moved around in one pile before it is equal to the size and shape of the other.

The EMD can be formally described as:

$$EMD(A,B) = \frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{A,B}\, d_{A,B}}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{A,B}},$$

where *A* and *B* are two distinct distributions, $f_{a,b}$ is the flow of mass between A and B, and $d_{a,b}$ is the penalization applied due to the distance traveled between A and B based on a defined function *D*. The EMD is the minimal amount of mass moved across a minimum distance and is normalized by the total flow between the two distributions (denominator).

In our application, we demonstrate the EMD's suitability to study the statistical difference in the mortality distributions (life table $d_x$) of two populations, and define it as thus:

$$EMD(d_A, d_B) = \frac{\sum_{i=0}^{\infty} \sum_{j=0}^{\infty} f_{i,j} p_{i,j}}{\sum_{i=0}^{\infty} \sum_{j=0}^{\infty} f_{i,j}},$$

where *i* and *j* represent a location along the distinct $d_{(x,t)}$ distributions *A* and *B*, $f_{i,j}$ is the flow between *i* in distribution *A* and *j* in distribution *B*, *p* is a penalization matrix of distance traveled between the points.

Solving the EMD can be done in different ways. In the special one-dimensional case, the EMD can simply be computed as the absolute difference in the cumulative distribution functions (CDF) of the two distributions:

$$EMD(A, B) = \int_{-\infty}^{\infty} |F_A(x) - F_B(x)| dx,$$

where $F_A$ and $F_B$ are the CDFs for the distributions in question. This number is often also normalized by the total value of the distribution, if it is not overwise equal to one. The Hungarian algorithm can also be used to solve the EMD in a two-dimensional approach (Kuhn, 2004). In these and in higher-dimensional instances, the EMD may be solved using other algorithms that minimize the cost (distance penalization) versus flow ("dirt" moved) within the context of the transportation problem.

**Data and application**

We calculate the 1D EMD from age-specific population and all-cause death counts at the NUTS-3 or NUTS-3 level based on official mortality data routinely collected by national statistical offices, which we harmonize across time and space to reflect administrative changes. This data compromises 526 subnational regions from 14 countries in Europe from which we calculate annual period life tables from 2001-2021 (Wilmoth et al, 2021). Using a Eurostat provided multi-polygon shapefile (EUROSTAT), we identify 1387 pairwise first-order relationships between adjacent regions (i.e. pairs of regions that border each other) reflecting 179 international and 1073 domestic borders for which we calculate the EMD.

To demonstrate an application of the 2D EMD, we focus on 38 regions (11 international and 37 domestic borders) within and between Austria, Czechia, and, Slovakia. After deriving multiple-decrement life tables for each region with five distinct causes (Lung Cancer, Other Cancer, CVD, External Causes, and Other), we calculate the 2D EMD for each two bordering regions considering the between-region dissimilarities in the distribution of *age* and *cause of death* (our 1D application focuses on distributional differences in *age*).

**Selected results**

Figure 1 show the results of the 1D EMD for our subregions of focus across Europe (women) during two time periods; lighter colors represent more similar lifetable age at death distributions, while darker colors represent larger distance between these distributions. Most stark, a strong mortality border persists across the former Iron Curtain in both time periods, but other cultural and language borders (for example, the French/Flemish speaking regions of Belgium) also exist. Figure 2 and the tables within it highlight the higher and wider distribution of EMDs along international vs. domestic borders, suggesting wider variability between countries versus within them.

**Summary and Outlook**

We introduce the Earth Mover's Distance and optimal transport theory and a way to measure dissimilarity between mortality distributions using statistical distance. The EMD simplifies the difference between two populations into a single number, and can consider additional stratification within these populations, thereby including multidimensional facets of death distributions. We show an application of the EMD to quantify

differences in mortality between subregions in Europe using age and cause of death, highlighting nuance in mortality differences between regions and the relative importance of international borders additionally as mortality boundaries.

In our examples using mortality distributions, the EMD is mostly limited to the lifetable context, as it requires two distributions of equal size (or standardized) and dimensionality. Moreover, the EMD calculates only the distance between the distributions, and therefore in our application does not infer which region has higher mortality than another. However, because the EMD can be calculated for any two standardized distributions, its possible applications in demography are limited to neither cause of death distributions nor mortality; for example, education level or occupation data could be used as stratification measures when comparing fertility distributions between two populations. We advocate for the use of the EMD across population research to consider complex data when understanding differences between groups.

**Figures**

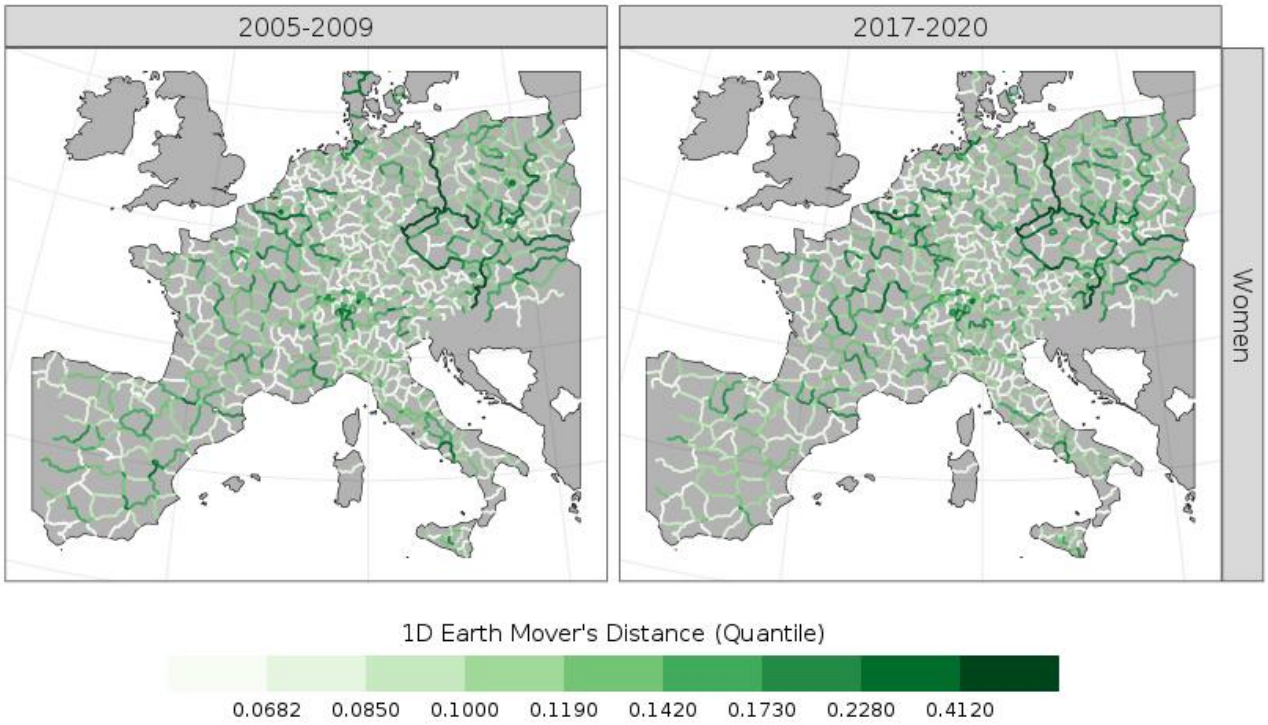


*Figure 1: One dimensional EMD in 2005-09 and 2017-20 for women. Each line is colored according to the EMD for the two lifetable age-at-death distributions for the two bordering regions.*

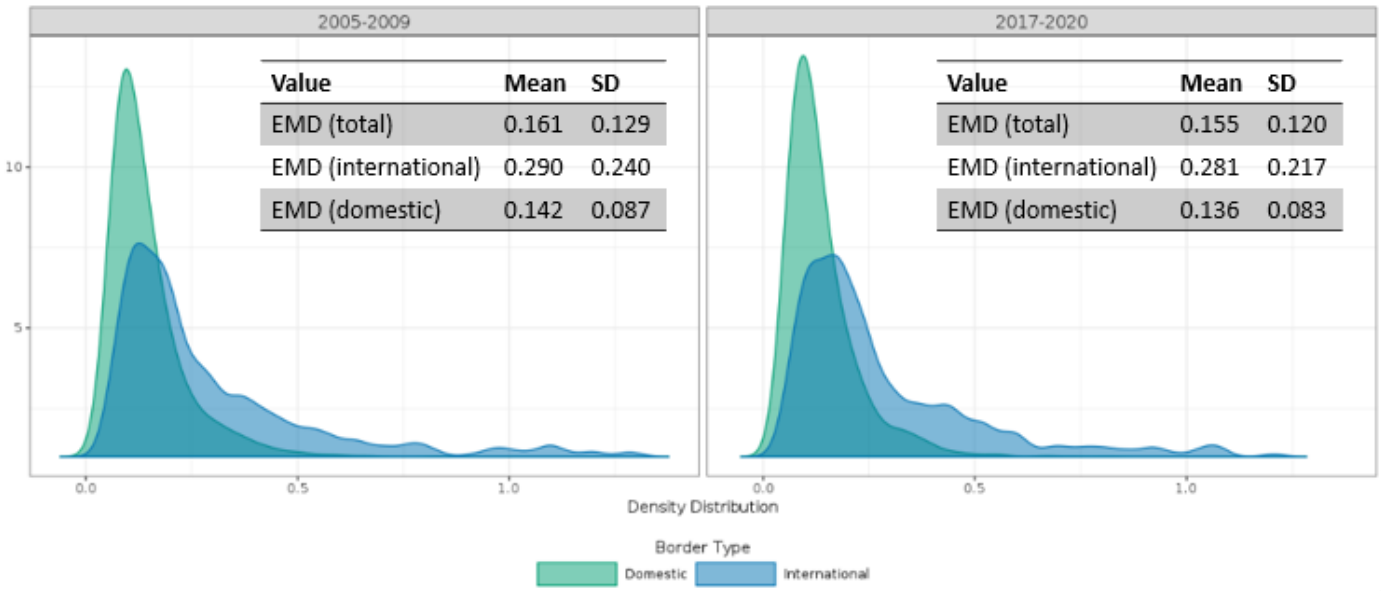


*Figure 2: Distribution of the 1-D EMD in 2005-09 and 2017-20, grouped by international and domestic borders. Tables show mean and standard deviations of these values.*

## References


Bonnet, F. and d'Albis, H. (2020), Spatial Inequality in Mortality in France over the Past Two Centuries. Population and Development Review, 46: 145–168.

Bramajo, Octavio, Iñaki Permanyer, and Amand Blanes. 2023. Regional inequalities in life expectancy and lifespan variation by educational attainment in Spain, 2014–2018. Population, Space and Place 29(e2628).

Eurostat. Territorial units for statistics (NUTS) - Eurostat. (2022). https://ec.europa.eu/eurostat/de/web/gisco/geodata/statistical-units/territorial-units-statistics

Hrzic, Rok, Tobias Vogt, Helmut Brand and Fanny Janssen. 2023. District-level mortality convergence in reunified Germany: long-term trends and contextual determinants. Demography 60(1): 303–325.

Klüsener, Sebastian, Brienna Perelli-Harris, and Nora Sanchez Gassen. 2013. Spatial Aspects of the Rise of Nonmarital Fertility Across Europe Since 1960: The Role of States and Regions in Shaping Patterns of Change / Aspects spatiaux de l'augmentation de la fécondité hors mariage en Europe depuis 1960: Le rôle des États et des régions dans l'élaboration de modèles de transformation. European Journal of Population / Revue Européenne de Démographie, 29(2), 137–165.

Kuhn, H. W. (2004). The Hungarian method for the assignment problem. Naval Research Logistics (NRL), 52(1), 7-21.

Monge, G. Memoire sur la Theorie des D'eblais et des Remblais. Histoire de l'Acad. des Sciences de Paris, 1781.

Mühlichen, Michael, Mathias Lerch, Markus Sauerberg, Pavel Grigoriev. 2023. Different health systems – Different mortality outcomes? Regional disparities in avoidable mortality across German-speaking Europe, 1992–2019. Social Science & Medicine 329: 115976.

Rau, Roland, and Carl P Schmertmann. 2020. District-level life expectancy in Germany. Deutsches Ärzteblatt International 117: 493–9.

Richardson, Elizabeth, Jamie Pearce, Richard Mitchell, Niamh K. Shortt, and Helena Tunstall. 2013. Have regional inequalities in life expectancy widened within the European Union between 1991 and 2008? European Journal of Public Health 24(3): 357-363.

Rubner, Yossi, Carlo Tomasi, and Leonidas J. Guibas. 2000. The earth mover's distance as a metric for image retrieval. International journal of computer vision, 40:99-121.

Schootman, M., Chien, L., Yun, S., & Pruitt, S. L. 2016. Explaining large mortality differences between adjacent counties: a cross-sectional study. BMC public health, 16:681.

Vallender, S. S. (1974). Calculation of the Wasserstein distance between probability distributions on the line. Theory of Probability & Its Applications, 18(4), 784-786.

Wilmoth, J. R., Andreev, K., Jdanov, D., Glei, D. A., Boe, C., Bubenheim, M., ... & Vachon, P. (2021). Methods protocol for the human mortality database. University of California, Berkeley, and Max Planck Institute for Demographic Research, Rostock. URL: http://mortality. org [version 26/01/2021].