## Income disparities in (multi)morbidity in premature death: retrospective analysis of Danish registry data

### Description of the topic

Premature deaths are often attributed to socioeconomic inequalities. For example, among the 2018–2019 Danish death cohort, individuals who died at ages 80–85 had about 10% higher inflation-adjusted income at age 60 than those who died at ages 65–70. Despite different income levels, income variation relative to the median is very similar in both groups. Thus, those who die prematurely are comparably diverse in income to those who die at older ages. Does the socioeconomic gradient in health disparities persist within the group of premature deaths?

This paper addresses that question. We conduct a retrospective analysis of (multi)morbidity for the 2018–2019 death cohort aged 60–85 in Denmark, reconstructing health trajectories back to 2000 to examine (multi)morbidity disparities by income. We define two income groups: (i) individuals who remained below the median income from 2000 until death, and (ii) individuals who remained above the median during the same period.

### Theoretical focus

Pro-rich gaps almost inevitably emerge across the spectrum of epidemiological and public health domains (Mair & Jani, 2020; Sigglekow et al., 2023; Jensen et al., 2017; Dijkstra & Horstman, 2021). Beginning with the most fundamental aspect of human life—its length—individuals with higher socioeconomic status (SES) tend to outlive those with lower SES (Marmot et al., 1987). This pattern, although varying in magnitude, is consistently observed across all low-mortality countries (Berkman, 2009), regardless of how egalitarian they are (Kunst et al., 2005). However, disparities in lifespan are merely the final reflection of health inequalities that develop long before death. Research has demonstrated that lower SES is associated with an earlier onset of disease-related health decline (Alvarez-Galvez et al., 2023) and a higher burden of multimorbidity (Mair & Jani, 2020; Jensen et al., 2017; Pathirana & Jackson, 2018; Froelich et al., 2019). This pattern holds irrespective of how SES is operationalized: whether through household income, wealth, educational attainment, occupation, parental socioeconomic characteristics, area-level deprivation, poverty indices, or other measures—all lead to similar conclusions about the direction of the associations (Alfonzo et al., 2022).

Hence, extensive work has been conducted on SES-related inequalities in morbidity so far. However, the majority of this work is one-sided, primarily due to its reliance on cross-sectional data and an exclusive focus on the elderly population (Tan et al., 2024; Alfonzo et al., 2022; Ferry et al., 2022; Skou et al., 2022). Consequently, this approach may limit the understanding of health disparities across socioeconomic groups for the following reasons. The multimorbid elderly population with low SES may not accurately reflect the experience of average low SES population but instead represents a highly selective sample of individuals who have been able to resist the higher risks of morbidity and associated mortality earlier in life, which are typical for their socioeconomic group. In other words, studies based on cross-sectional data of the elderly are more likely to include less frail individuals within low SES groups. Such selection is unlikely to apply to high SES groups, where multimorbidity develops on average 10-15 years later (Skou et al., 2022; Alvarez- Galvez et al., 2023), and individuals typically have a life expectancy that is around 2 to 17 years longer (Murtin et al., 2017), based on data from OECD countries and using educational attainment as a proxy for SES.

Moreover, a recent review of the socioeconomic determinants of multimorbidity also highlighted that existing studies work with rather limited lists of conditions (Alfonzo et al., 2022), with the median number of diseases considered being 16 (IQR 9-27).

In this paper, we aim to avoid the previously described stereotypes and pose the following question: **How do inequalities in multimorbidity by income manifest among individuals who, despite their SES, died around the same time and prematurely between the ages of 60 and 85?** Thus, we focus on a population that, regardless of SES, is disadvantaged at the end of life because they die before the modal age at death. Moreover, we remain at the 3-digit ICD-10 code level resulting in an analysis that encompasses approximately 2,000 distinct diseases.

**Data**

This population-wide study integrates data from the Danish Death Registry (DDR), the Danish National Patient Registry (LPR), and the Danish Income Registry (DIR). From the DDR, we identified our study population based on year of death (2018/2019) and age at death (60–85, N=59,985). These selected individuals were subsequently traced in the LPR to retrieve all primary and secondary diagnoses they received during 2000–2018/2019. The analysis treats all diagnoses equally, without distinguishing between primary and secondary diagnoses. Another data extracted from the LPR were the dates of diagnosis. To determine the chronological order of diseases, we only used the year of diagnosis. Finally, to determine the income level of individuals, we used the equivalized disposable household income, which is a gold standard that is recommended to be used when analyzing income factors using Danish registries (Hjorth et al, 2025).

**Methods**

We consider two perspectives on the associations between diseases: (i) co-occurrence and (ii) transitions. The former refers to pairs of diseases that appear within the same patient during the same calendar year. The latter pertains to pairs of diseases that occur sequentially, with one following the other, thus representing ordered pairs. The procedure for constructing these disease pairs is illustrated in the following example:

For example, a person diagnosed at time T with diseases A, B, C, and B would have the following disease pairs: AB, AC, and BC. Repeated diagnoses of the same disease, which occur when aggregating from 4-digit ICD-10 codes to 3-digit codes, were not taken into account, and the sequence of disease diagnoses within the same year was not distinguished. If the same individual is diagnosed with diseases A, C, and D at year T + n, then the identified pairs of diseases between years T and T + n are: AC, AD, BA, BC, BD, CA, CD, AA, and CC. When analyzing transitions across different time periods, we also include "remaining with diagnosis," which includes pairs of the same diseases, as AA and CC in this example. However, redundant pairs are not considered.

At the end of the data processing, we obtained a system of block matrices $M_{T:T+n}$where T is the first year of the analysis and $n \in \langle 0\,;19 \rangle$:

$$\begin{bmatrix} M_{T:T} & M_{T:T+1} & \cdots & M_{T:T+n} \\ 0 & M_{T+1:T+1} & \cdots & M_{T+1:T+n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & M_{T+n:T+n} \end{bmatrix},$$

where each block matrix $M_{T:T}, M_{T:T+1}, \ldots . M_{T+n:T+n}$ consist of count of disease pairs occurring simultaneously:

$$M_{T+n:T+n} = \begin{bmatrix} 0 & n_{ij} & n_{ik} \\ n_{ji} & 0 & n_{jk} \\ n_{ki} & n_{kj} & 0 \end{bmatrix}$$

Or occurring subsequently, indicating transitions between diagnosis:

$$M_{T+n:T+m} = \begin{bmatrix} n_{ii} & n_{ij} & n_{ik} \\ n_{ji} & n_{jj} & n_{jk} \\ n_{ki} & n_{kj} & n_{kk} \end{bmatrix}$$

Where $m \in \langle 0\,;19 \rangle \wedge m \neq n.$

To later derive the indicators, we calculated, for each block of the matrix $M_{T:T+n}$, the 2×2 contingency table with death counts, using the following notation:

| | | Disease j | | |
|---|---|---|---|---|
| | | 1 | 0 | |
| Disease i | 1 | $n_{ij}$ | $n_{i\neg j}$ | $n_i$ |
| | 0 | $n_{\neg ij}$ | $n_{\neg i\neg j}$ | $n_{\neg i}$ |
| | | $n_j$ | $n_{\neg j}$ | n |

¬ = does not have

Where n was total number of unique IDs at time T out of the study population.

The data processing was conducted separately for income groups. The population was divided into two income groups based on the median income to avoid excessive granularity in the analysis. Naturally, both income levels and their nominal values change over time. Moreover, income levels may fluctuate due to transitions between morbid states or the accumulation of health conditions. To minimize this bias, individuals who transitioned between lower and higher SES groups during 2000–2019—defined by time-varying median income—were excluded from the analysis. This exclusion is also theoretically justified, as we consider income stability to be an additional indicator of financial advantage. As a result, the sample size reduced to 34,631 individuals. These individuals were split into two income groups—low and high—based on their income relative to the median for each year from 2000 to 2018/19. The low-income group (LIG) included 17,393 people, and the high-income group (HIG) included 17,238 people.

We conducted the analysis on 3-digit level of ICD-10 codes. All diagnoses were included, except for Z* codes, which indicate contact with medical services. We restricted the analysis to co-occurrences and transitions that occurred at least 50 times. Overall, the analysis encompassed 3,919 distinct transitions between 238 diseases and 2,012 unique co-occurrences among 196 diseases.

For the analysis of the strength of association between **co-occurring diseases**, we use odds ratios (OR), as usually used in similar studies (Chmiel et al., 2014, Hidalgo et al., 2009; Fotouhi et al., 2018;

Kim et al., 2016; Khan, 2018). The well-known formula for the odds ratio, expressed with the notation shown above, is as follows (Schmidt & Kohlmann, 2008):

$$¿ = \frac{\frac{N_{ij}}{N_{\neg ij}}}{\frac{N_{i\neg j}}{N_{\neg i\neg j}}}$$

For the analysis of **associations across time**, we use the risk increase (RI) indicator, which quantifies how much the presence of the first disease increases the likelihood of developing the second disease, relative to its probability in the general population. The formula for the risk increase indicator can be expressed as follows (Kerexeta-Sarriegi et al., 2024):

$$RI = \frac{N_{ij}}{N_i} \times \frac{N}{N_j}$$

Inequalities between income groups are evaluated using the index of RI, comparing individuals with lower and higher income, with the higher-income group serving as the reference. Values of the RI index above 1 indicate a disadvantage for the lower-income group, reflecting an increased risk of developing disease associations over time.

**(Expected) findings**

First, we explored the relationship between age at death, years prior to death (remaining years of life), total number of diseases, and income level (Figure 1). Panel A displays notable differences in age at premature deaths across income groups. Considering premature death only, the HIG tend to die on average around 7 years younger than LIG.

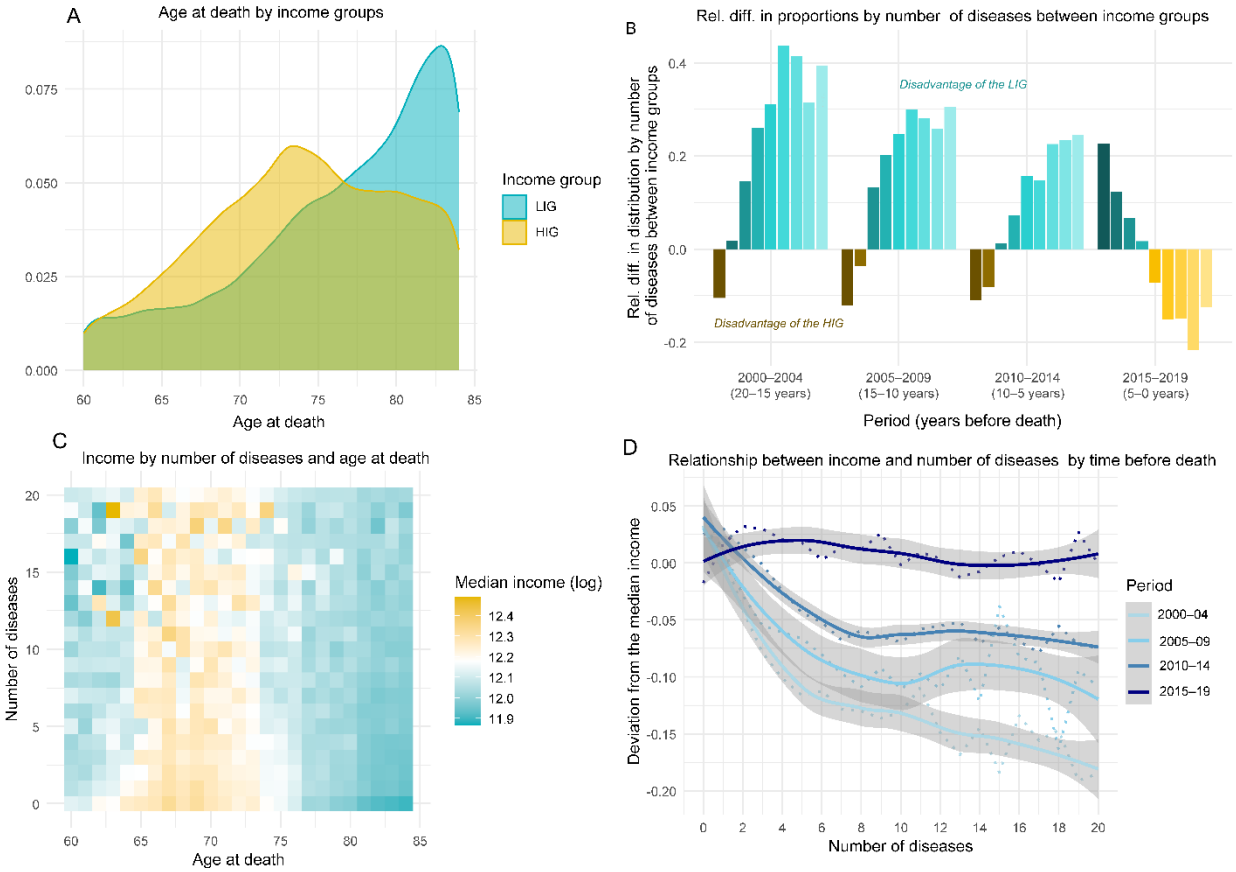
Panel B depicts the relative differences in the distribution of the number of diseases between LIG and HIG. Negative values indicate a higher relative frequency of HIG with a specific number of diseases, while positive values suggest a higher frequency of LIG. As individuals approach death, the relative dominance of LIG among those with complex multimorbidity diminishes. This trend reverses in the last quinquennial, when individuals with complex multimorbidity dominate within HIG, whereas the proportions of people with fewer than three diseases become higher in LIG than in HIG.

Panel C shows income levels by the number of diseases and age at death. The color progression along the horizontal axis highlights that the premature death of individuals in the HIG usually occurs between 65 and 75 years of age, as already evidenced on Panel A. In contrast, premature deaths in the LIG tend to happen later in life, often with no or only a few diseases. Notably, individuals with complex multimorbidity (>10 diseases) who die before age 70 are a highly diverse group in terms of socioeconomic status.

Finally, Panel D contrasts the remaining years of life (period with income and the total number of diagnoses. Income is expressed as deviations from the median within each period to account for inflation. The gradient of increasing income disadvantage with the increasing number of diseases tends to diminish with the progression towards death. In the last five years before premature death, individuals with more diseases no longer fall behind the income norm, meaning they no longer exhibit negative income deviations.

To sum up, descriptive findings indicate that among premature deaths, HIG likely experience more pronounced negative outcomes as death approaches: their age at death is lower compared to LIG, and multimorbidity increases. Additionally, shortly before premature death, income disparities by total number of diseases tend to narrow.

*Figure 1: Premature death by age at death, years prior to death (remaining years of life), total number of diseases, and income level*




Next, we conducted an analysis of disparities in specific combinations of diseases, which revealed that the most commonly treated conditions at the population level—such as diabetes mellitus (E11), atrial fibrillation (I48), heart failure (I50), and COPD (J44)—often pose a higher risk of disease accumulation in the HIG. This is especially evident when comparing the periods 2000-2004 with 2005-2009, and 2010-2014 with 2015-2019. Our analysis of odds ratios showed that, long before death, the LIG generally exhibits stronger associations between chronic diseases. However, this pattern reverses shortly before death. As a next step, we plan to examine disparities in specific disease combinations more thoroughly.

In summary, we employed an unconventional approach to analyze SES inequalities in multimorbidity. Preliminary results show that among premature deaths, the HIG appears to be at a disadvantage compared to LIG, showing earlier mortality, a higher multidisease burden, and increased susceptibility to future disease accumulation. These results should be interpreted in the context of differing rates of healthcare engagement between HIG and LIG, among the other.

**References**


Alvarez-Galvez, J., Ortega-Martin, E., Carretero-Bravo, J., Perez-Munoz, C., Suarez-Lledo, V., & Ramos-Fiol, B. (2023). Social determinants of multimorbidity patterns: A systematic review. Frontiers in Public Health, 11, 1081518.

Berkman, L. F. (2009). Social epidemiology: social determinants of health in the United States: are we losing ground?. Annual review of public health, 30(1), 27-41.

Dijkstra, I., & Horstman, K. (2021). 'Known to be unhealthy': Exploring how social epidemiological research constructs the category of low socioeconomic status. Social Science & Medicine, 285, 114263.

Ferry, F. R., Rosato, M. G., Curran, E. J., O'Reilly, D., & Leavey, G. (2022). Multimorbidity among persons aged 25–64 years: a population-based study of social determinants and all-cause mortality. Journal of Public Health, 44(1), e59-e67.

Fotouhi, B., Momeni, N., Riolo, M. A., & Buckeridge, D. L. (2018). Statistical methods for constructing disease comorbidity networks from longitudinal inpatient data. *Applied Network Science*, *3*(1), 46.

Frølich, A., Ghith, N., Schiøtz, M., Jacobsen, R., & Stockmarr, A. (2019). Multimorbidity, healthcare utilization and socioeconomic status: a register-based study in Denmark. PloS one, 14(8), e0214183.

Hidalgo, C. A., Blumm, N., Barabási, A. L., & Christakis, N. A. (2009). A dynamic network approach for the study of human phenotypes. *PLoS computational biology*, *5*(4), e1000353.

Hjorth, C. F., Kjærulff, T. M., Thomsen, M. K., Cronin-Fenton, D., Dalton, S. O., & Olsen, M. H. (2025). SEPLINE: Socioeconomic Position in Epidemiological Research—A National Guideline on Danish Registry Data. *Clinical Epidemiology*, 593-624.

Chmiel, A., Klimek, P., & Thurner, S. (2014). Spreading of diseases through comorbidity networks across life and gender. *New Journal of Physics*, *16*(11), 115013.

Kerexeta-Sarriegi, J., García-Navarro, T., Rollan-Martinez-Herrera, M., Larburu, N., Espejo-Mambié, M. D., Iraola, A. B., & Graña, M. (2024). Analysing disease trajectories in a cohort of 71,849 Patients: A visual analytics and statistical approach. *International Journal of Medical Informatics*, *188*, 105466.

Khan, A., Uddin, S., & Srinivasan, U. (2018). Comorbidity network for chronic disease: A novel approach to understand type 2 diabetes progression. *International journal of medical informatics*, *115*, 1-9.

Kim, J. H., Son, K. Y., Shin, D. W., Kim, S. H., Yun, J. W., Shin, J. H., ... & Yun, J. M. (2016). Network analysis of human diseases using Korean nationwide claims data. *Journal of biomedical informatics*, *61*, 276-282.

Kunst, A. E., Bos, V., Lahelma, E., Bartley, M., Lissau, I., Regidor, E., ... & Mackenbach, J. P. (2005). Trends in socioeconomic inequalities in self-assessed health in 10 European countries. International journal of epidemiology, 34(2), 295-305.

Lund Jensen, N., Pedersen, H. S., Vestergaard, M., Mercer, S. W., Glümer, C., & Prior, A. (2017). The impact of socioeconomic status and multimorbidity on mortality: a population-based cohort study. Clinical Epidemiology, 9, 279–289. https://doi.org/10.2147/CLEP.S129415

Mair, F. S., & Jani, B. D. (2020). Emerging trends and future research on the role of socioeconomic status in chronic illness and multimorbidity. The Lancet Public Health, 5(3), e128-e129.

Marmot, M. G., Kogevinas, M., & Elston, M. A. (1987). Social/economic status and disease. Annual review of public health, 8(1), 111-135.