MINISTERIO
DE ECONOMÍA, INDUSTRIA
Y COMPETITIVIDAD

CSIC
CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS

CCHS
Centro de Ciencias
Humanas y Sociales
Instituto de Economía, Geografía y Demografía

LONGPOP

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 676060

Methodologies and Data mining techniques for the analysis of Big Data based on Longitudinal Population and Epidemiological Registers

IRP 10.1. Data model for the combined analysis of inequalities in health with different population registers

Mathias Voigt

Early-stage Researcher

Disclaimer: This publication reflects only the author's view and the agency is not responsible for any use that may be made of the information it contains.

Introduction

There is evidence in all regions of the world, even in most high-income countries with comprehensive and affordable health care systems in place, which suggests that health and survival are strongly associated with income, wealth, education and type of occupation. Numerous analyses confirm for instance the existence of substantial mortality advantages for better educated and wealthier individuals over others [1–6]. Particularly in times of increasing social inequalities in all OECD countries [7–9], the association between social position and structural inequalities with respect to health measures leads to concerning conclusions for future population health [10]. An increasing body of research is dedicated to explaining and projecting such health inequalities and in turn leads to an increasing need for unified measurement strategies and comparability. Aiming to fill part of this void, we introduce a data model which, in each setting, combines information from various data sources in order to assess persisting or prospective health inequalities within a population. Given the growing accessibility and use of administratively produced data, the model is specifically adapted to population register data. Examples are based on data sources covering the Spanish population but not limited to it. This report on the data model for combined analysis of inequalities in health is structured in three main parts. In the first part, the occurrence and development of structural social inequalities as most prominent but least desirable underlying cause for health inequity are described. We try to explain how different socioeconomic aspects can be related to health on an individual and population level. The second part is centered around health as central concept of analysis of health inequalities. We aim to highlight the potential pitfalls of working with this highly dimensional and shifting concept in relation to population science. The concluding part is a description of our data model as an attempt to structure the assessment of health inequalities with population registers.

Measuring social inequalities to predict health

There are countless possible pathways to explain why individuals or groups are healthier than others. The largest part of the occurring disparities at an individual level can probably be traced back to variations in genetic predisposition, the involvement in accidents, and behavioral aspects regarding diet, smoking, and risk adversity. From the perspective of public health research, however, a simple aggregation of these measures is not sufficient for the comprehension of patterns and structural differences which need to be understood to provide large-scale interventions. Apart from the direct effects we need to ask how underlying factors

lead to the aforementioned patterns which often emerge based on existing socioeconomic differences. Wealthier and higher educated people tend on average to be healthier and can expect to live longer than their less wealthy and less educated counterparts. This indirect association between the socioeconomic position and the physical and mental wellbeing can be understood as the central concern of the research on health inequalities. Moreover, understanding and analyzing this relationship seems to be crucial regarding predictions and action on various social spheres, including the assurance of a certain degree of social fairness and the efficiency of public health interventions.

Although health and socioeconomic position can be related within a theoretical framework as, for instance, by the absolute and relative income hypothesis [see 11, 12], quantifying and comparing structural social inequalities often requires a sophisticated operationalization to avoid numerous pitfalls [8]. On the one hand, researchers face the problem that data on socioeconomic position is time dependent on at least two different scales. While temporally varying macroeconomic developments like recessions or boom periods affect price levels, salaries, and labor market demands, the occurrence and timing of particular life course events, like the birth of a child or the completion of a degree, has substantial impact on the socioeconomic position of individuals and is interrelated with the macro-level development. In order to make general statements or well-founded predictions, such dynamic requires the availability of time series data or retrospective data. Given that the individual socioeconomic status is not static in time, it might be necessary to account for biographical trends and common life course trajectories. Although there appear to be geographical differences, two macro hypotheses have emerged which attempt to broadly describe and generalize trajectories of social inequalities and structural disparities in health over the life course. While the accumulation theory suggests that members of lower socioeconomic groups in contrast to more affluent and wealthier individuals accumulate relative disadvantages over time, the convergence-divergence hypothesis argues that socioeconomically based health inequalities might peak at early old age before they are balanced by social security and retirement payments [13, 14].

If time series and individual level information are available on the other hand, one still faces the challenge of the relative arbitrary decision of choosing a set of threshold values. Therefore, it will be necessary to decide how to define who is poor and who is not? Furthermore, it requires to make the rather difficult distinction between inequality which might be desirable

as a kind of economic incentive and inequality which is the result of structural disadvantages. Making these decisions is often not easy and requires a careful assessment of different aspects of validity. Regarding health inequalities, the big theoretical question is centered around the possible outcomes of experienced social inequality on an individuals' or a groups health, often over substantial periods of time. Consequently, social inequalities functions as explanatory concept for existing inequalities in various health and mortality measures. On the other hand, it can be argued that the state of health determines your access to education and the job market, an endogenous relationship which has be controlled for if one assumes that health inequalities emerge from social inequalities [15].

Health – Assessing a multi-factorial concept

The quantitative research on inequalities in health and mortality plays an important role in the understanding and quantification of macro developments which not only address the aspects of social fairness but support policy planning in fields like the estimation of the future care load or urban planning of health facilities. While estimating and measuring mortality is the bread and butter of demographers, health can often only be measured through absence, or in other words the occurrence of a health problem, or in another indirect fashion. It is surprisingly difficult to decide between healthy and unhealthy in the first place. While self-assessed general well-being is often used as a relatively economical approximate the overall health status of an individual, the application of newer and more objective measures revealed that differences in self-assessed health seem to exaggerate actual inequalities and regional disparities within and between countries [16, 17].

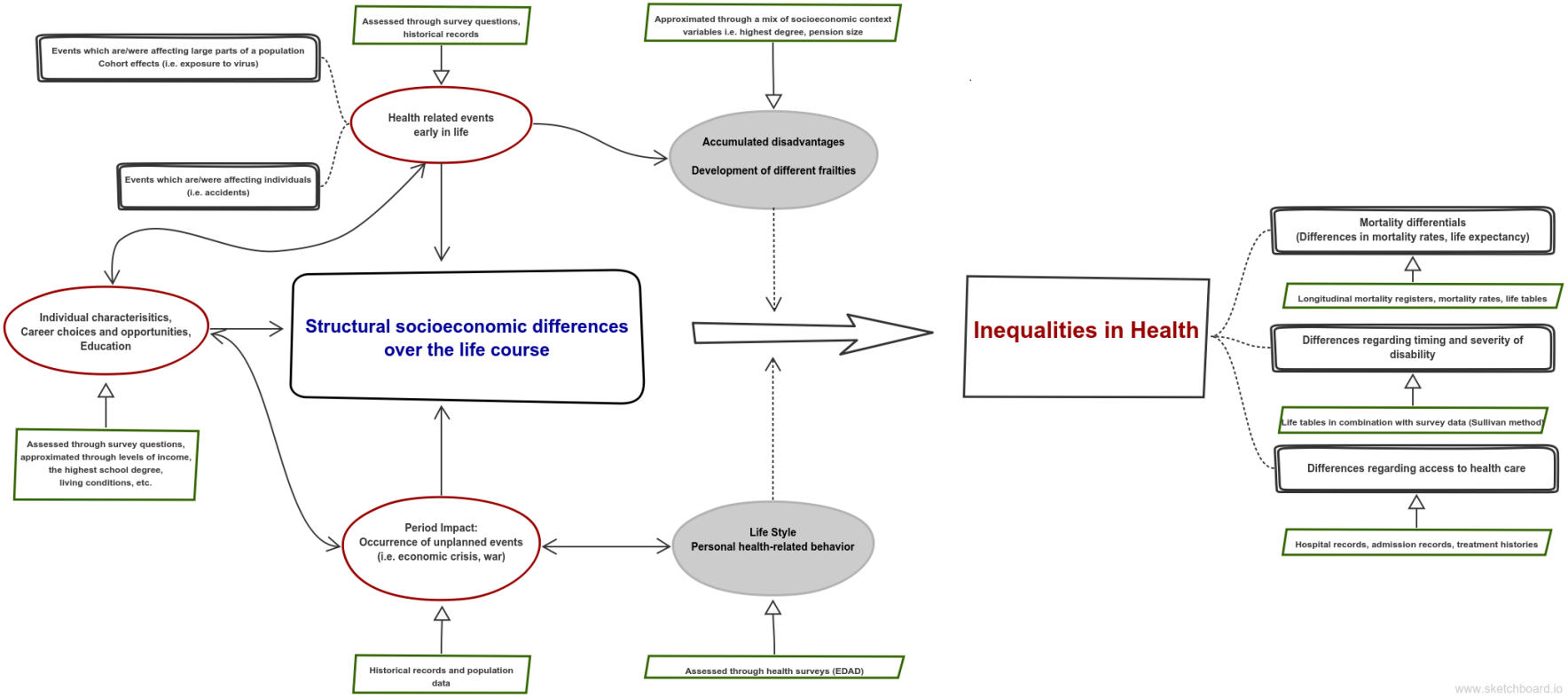
Health measures which are aggregated on a population level like incidence or prevalence rates can be obtained from various open data sources including the WHO website, the most accurate way to assess socioeconomically based inequalities in health status is arguably to apply individual, longitudinal data. It might not be the only way but these data type allows for disease trajectories and a linkage to socioeconomic information. Even if we assume such data is available, there are further challenges ahead the road for the research on health inequalities. Among those are the lack of medical knowledge, missing information on co-morbidity, and the sensitivity of individual health data [18].

Nevertheless, there is guidance and several useful tools like the International Classification of Diseases (ICD), an internationally accepted guideline for coding and grouping morbidity and mortality statistics, which eleventh revision is about to be released soon [19]. The increasing accessibility to administratively collected morbidity data and the linking of various register-based data bases further increases the possibilities to estimate inequalities in health and the association with socioeconomic disadvantages.

The data model for the combined analysis of health inequalities

To analyze the association between two highly dimensional concepts like socioeconomic position and health inequalities, it will be necessary to consider and discover various allies which potentially contribute to the explanation of a macro-situation. In an attempt to account for the above-mentioned pitfalls and difficulties, we developed a data model for the combined analysis of health inequalities with different population registers based on an example applied to a register-based longitudinal population dataset, the Base de Datos Longitudinal de Población de Andalucía (BDLPA). This highly integrated, relational data infrastructure, provided by the Institute of Statistics and Cartography of Andalusia, probably offers more opportunities for assessing different areas of population health and environmental impact than most available data sets. This will, however, offer us the possibility to create a more complex and adaptable model which can then be reduced based on the available data structure of a prospective user. If particular information cannot be assessed, one can use a related, parallel ally to link the two constructs or analyze a part of the underlying association. We are aware what the model, depicted in figure 1 in the appendix, can do. It does not give a universal answer to the, as above mentioned, highly complex relationship of the two multi-dimensional constructs. Our data model can be rather understood as a suggestion how to approach the analysis of health inequalities given the availability of data sources, which are in the best case linkable.

Accounting for time trends and trajectories is essential for individual level assessment of such an association. Health problems as well as socioeconomic disadvantages can accumulate over a life time. At the same time, there are several hypotheses which stress the importance of early life events and therefore put more weight on this period.


Assessed through survey questions, historical records
Approximated through a mix of socioeconomic context variables i.e. highest degree, pension size
Events which are/were affecting large parts of a population Cohort effects (i.e. exposure to virus)
Health related events early in life
Accumulated disadvantages
Development of different frailties
Events which are/were affecting individuals (i.e. accidents)
Mortality differentials (Differences in mortality rates, life expectancy)
Longitudinal mortality registers, mortality rates, life tables
Individual characterisitics, Career choices and opportunities, Education
Structural socioeconomic differences over the life course
Inequalities in Health
Differences regarding timing and severity of disability
Life tables in combination with survey data (Sullivan method)
Assessed through survey questions, approximated through levels of income, the highest school degree, living conditions, etc.
Differences regarding access to health care
Period Impact: Occurrence of unplanned events (i.e. economic crisis, war)
Life Style Personal health-related behavior
Hospital records, admission records, treatment histories
Historical records and population data
Assessed through health surveys (EDAD)
www.sketchboard.io

References

[1] J. P. Mackenbach, M. Bopp, P. Deboosere, K. Kovacs, M. Leinsalu, P. Martikainen, G. Menvielle, E. Regidor, and R. de Gelder, “Determinants of the magnitude of socioeconomic inequalities in mortality: A study of 17 European countries,” Health & Place, vol. 47, no. Supplement C, pp. 44–53, 2017.

[2] K. G. Manton, E. Stallard, and L. Corder, “Education-specific estimates of life expectancy and age-specific disability in the U.S. elderly population 1982 to 1991,” Journal of Aging and Health, vol. 9, no. 4, pp. 419–450, 1997.

[3] J. K. Montez, R. A. Hummer, M. D. Hayward, H. Woo, and R. G. Rogers, “Trends in the educational gradient of U.S. adult mortality from 1986 through 2006 by race, gender, and age group,” Research on Aging, vol. 33, no. 2, pp. 145–171, 2011.

[4] S. J. Olshansky, T. Antonucci, L. Berkman, R. H. Binstock, A. Boersch-Supan, J. T. Cacioppo, B. A. Carnes, L. L. Carstensen, L. P. Fried, and D. P. Goldman, “Differences in life expectancy due to race and educational differences are widening, and many may not catch up,” Health Affairs, vol. 31, no. 8, pp. 1803–1813, 2012.

[5] J. P. Mackenbach, I. Stirbu, A.-J. R. Roskam, M. M. Schaap, G. Menvielle, M. Leinsalu, and A. E. Kunst,

“Socioeconomic inequalities in health in 22 European countries,” New England Journal of Medicine, vol. 358, no. 23, pp. 2468–2481, 2008.

[6] J. P. Mackenbach, M. Karanikolos, and M. McKee, “The unequal health of Europeans: successes and failures of policies,” The Lancet, vol. 381, no. 9872, pp. 1125–1134, 2013.

[7] P. Doerrenberg and A. Peichl, “The impact of redistributive policies on inequality in OECD countries”, Applied Economics, vol. 46, no. 17, pp. 2066–2086, 2014.

[8] M. Corak, “Income inequality, equality of opportunity, and intergenerational mobility,” The Journal of Economic Perspectives, vol. 27, no. 3, pp. 79–102, 2013.

[9] R. Bachmann, P. Bechara, and S. Schaffner, “Wage inequality and wage mobility in Europe,” Review of Income and Wealth, vol. 62, no. 1, pp. 181–197, 2016.

[10] M. Marmot, “Social determinants of health inequalities,” The Lancet, vol. 365, no. 9464, pp. 1099 – 1104, 2005.

[11] K. E. Pickett and R. G. Wilkinson, “Income inequality and health: A causal review,” Social Science & Medicine, vol. 128, pp. 316–326, 2015.

[12] K. Karlsdotter, J. J. Martín Martín, and M. P. López del Amo González, “Multilevel analysis of income, income inequalities and health in Spain,” Social Science & Medicine, vol. 74, no. 7, pp. 1099–1106, 2012.

[13] E. K. Pavalko and J. Caputo, “Social inequality and health across the life course,” American Behavioral Scientist, vol. 57, no. 8, pp. 1040–1056, 2013.

[14] S. G. Prus, “Age, ses, and health: A population level analysis of health inequalities over the lifecourse,” Sociology of health & illness, vol. 29, no. 2, pp. 275–296, 2007.

[15] N. Ziebarth, “Measurement of health, health inequality, and reporting heterogeneity,” Social Science & Medicine, vol. 71, no. 1, pp. 116–124, 2010.

[16] H. J ¨ urges, “True health vs response styles: exploring cross-country differences in self-reported health,” Health economics, vol. 16, no. 2, pp. 163–178, 2007.

[17] M. Baker, M. Stabile, and C. Deri, “What do self-reported, objective, measures of health measure?” Journal of human Resources, vol. 39, no. 4, pp. 1067–1093, 2004.

[18] W. C. Cockerham, Medical Sociology. JohnWiley & Sons, Ltd, 2014.

[19] “ICD-11 Revision Conference,” report, World Health Organization (WHO), 2016.