Social Epidemiology: The Challenges and Opportunities of Worldwide Data Consortia

Carlotta Galeone, Rossella Bonzi, Federica Turati, Claudio Pelucchi, Matteo Rota, and Carlo La Vecchia

Abstract Over the last few decades, social epidemiology has developed as a solid epidemiology branch, focusing on understanding how social experiences influence population health. At the same time, growing of collaborative and interdisciplinary research led to the proliferation of multi-institutional consortia, able to assess and quantify risk-disease associations of interest with a higher degree of accuracy, to explore subgroups of the population, and to investigate interactions between environmental, genetic, and socioeconomic factors. Increasing evidence shows that low Socioeconomic Position (SEP) is a strong determinant of morbidity and premature mortality from selected non-communicable diseases, including several cancers. Thus, an accurate quantification of the impact of SEP on cancer risk is of major importance to plan public health interventions for cancer incidence and socioeconomic disparities reduction. Large data consortia as the Stomach Cancer Pooling (StoP) Project and the International Head and Neck Cancer Epidemiology (INHANCE), in which the University of Milan is proactively involved, allowed investigators to address the effects of education and household income, the main SEP determinants, on gastric and head and neck cancer, respectively, confirming the existence of a strong association between low SEP and those major neoplasms.

C. Galeone (✉) · R. Bonzi · F. Turati · C. Pelucchi · C. La Vecchia
Department of Clinical Sciences and Community Health, University of Milan, Milan, Italy
e-mail: carlo.lavecchia@unimi.it

C. La Vecchia
e-mail: carlotta.galeone@gmail.com

M. Rota
Department of Molecular and Translational Medicine, University of Brescia, Brescia, Italy

P. Mariani and M. Zenga (eds.), *Data Science and Social Research II*,
Studies in Classification, Data Analysis, and Knowledge Organization,
https://doi.org/10.1007/978-3-030-51222-4_14

1 Introduction

Social epidemiology is a relatively recent branch of epidemiology that aims to understand how social factors affect population health (Honjo 2004). One of the most important examples of sociostructural factors in social epidemiology is the study of social class in relation to Non-Communicable—or chronic—Diseases (NDCs). NDCs are diseases of long duration and generally slow progression and are the leading causes of health issues worldwide, accounting for 63% of all annual deaths globally. Cardiovascular diseases, cancers, chronic respiratory diseases, and diabetes are the main types of NDCs, whose reduction, according to the World Health Organization (WHO) member states, in terms of mortality in people aged 30–70 has been committed to be achieved by the year 2025 (WHO 2018b; Bennett et al. 2018). NDCs are the result of a combination of genetic, physiological, environmental, and behavioral factors, exposing people to health, social, and economic challenges on a daily basis. Disadvantaged people are caught in a sort of vicious circle in which poverty causes illness and illness feeds poverty (Wagstaff 2002).

Consolidated evidence shows that NCDs morbidity and premature mortality is higher in low-income and middle-income countries, and, at least in high-income countries, in people with lower socioeconomic status, making NCDs an important obstacle to reducing global and national health inequalities (Marmot 2014; Wagstaff 2002; Niessen et al. 2018).

Socioeconomic Position (SEP) reflects the availability of cultural, material, and social resources that translate into advantages in terms of decision making, social network, lifestyle habits, and also access to health services. An accurate quantification of the impact of SEP on the risk of the disease is of major importance to plan public health interventions aimed to reduce NCDs incidence and socioeconomic disparities.

Among the most commonly diffused NCDs, cancer is the second leading cause of death worldwide, and responsible for an estimated 9.6 million deaths in 2018. Globally, about 1 in 6 deaths is due to cancer and approximately 70% of deaths from cancer overall occur in low- and middle-income countries (WHO 2018a).

A precise quantification of SEP impact on cancer is difficult to point out locally, as cancer incidence, survival, and mortality are subject to large variations across countries and, within countries, across social groups. Social inequalities, as well, are continuously evolving and reshaping over time as a reflection of the economic, political, social, legislative, and technological asset of the society. Differences among social groups have a strong impact on cancer at every stage of the disease by affecting the exposure to risk factors, and hence the likelihood to develop the disease, as well as the timely access to public health measures, diagnostic and treatment facilities and health-care services.

High-income countries show higher incidence rates of all cancers than most low- and middle-income countries, mainly because of environmental and lifestyle risk factors. At the same time, low- and middle-income countries often have similar or sometimes higher mortality rates from cancer than high-income countries, mainly because of a lack of access to timely diagnosis and proper treatment. However, within almost all countries, mortality rates for most cancer types are, to a disproportionate extent, higher for groups of the population with low socioeconomic position or otherwise disadvantaged, due to poorly designed health systems or limited or even inhibited access to preventive interventions, early detection, diagnosis, treatment, or/and palliative care (Vaccarella et al. 2019; Niessen et al. 2018).

Disparities in cancer care could be resolved if the highest achievable standards in health care were attained across countries at all economic levels. Failing to translate the excellent results constantly achieved in cancer knowledge through scientific research into effective action, in terms of health infrastructure and adequate basic services, still contributes to regional, national, and international health inequities.

The advent of collaborative and interdisciplinary research framework, along with the proliferation of multi-institutional research consortia during the last two decades, markedly affected cancer epidemiology. The National Cancer Institute's Epidemiology gives the most globally recognized definition of consortium as a 'group of scientists from multiple institutions who have agreed to participate in cooperative research efforts involving activities such as methods development and validation, pooling of information from more than one study for the purpose of combined analyses, and collaborative projects. Consortia are able to address scientific questions that cannot be addressed otherwise due to scope, resources, population size, or expertise. This cooperation usually involves multiple projects over an extended time period. Consortia can also be referred to as collaboratives (NIH 2019).

A general feature and strength of consortia is the easy and prompt communication among members for an interconnected sharing of study results. Such large collaborative groups benefit in terms of dissemination of research tools and information, from the establishment of web forums, public websites, or other global means of inter-diffuse communication. Quick and fluent knowledge and data sharing are on the basis for the coordination of the scientific research whose aim is to maximize the efficiency to understand, prevent, treat, and relieve the risk and hence the incidence of diseases on the population at a global level.

Moreover, the uniquely large data set on which consortia are based permit to define and quantify, with a degree of accuracy higher than ever before, the main effects of each risk factor of interest and to adequately address associations in subgroups of the population, as well as interaction between environmental, genetic, and socioeconomic factors.

2 Definition of Socioeconomic Position

SEP is a complex concept which involves several dimensions including education, work experience, and household income, access to material resources, prestige, and social position. All of these dimensions are associated, even though each of them accounts for different aspects of the socioeconomic stratification. In a broader sense, speaking of socioeconomic status involves referring to the most common forms of inequality (Geyer et al. 2006).

The assessment of socioeconomic position in the epidemiologic research is usually performed throughout the use of a series of indicators, traditionally education, occupation, and income, though their specific use often and strictly depends on data availability.

The strengths and limitations of the selected SEP determinants are herein briefly summarized, following a measure reliability order, starting from income, the less stable measure.

Individual, or, better, household income, which may be a useful indicator in particular for women or those who may not be the main earners in the household, reflects the material component of people everyday life. People with higher incomes are more likely to experience better living conditions, social services affordability, and healthy environment than lower income groups. However, income is the SEP indicator mostly subjected to changes, also on a short-term basis, it is age-dependent and it shows the highest non-response rate in epidemiological investigations when compared to other SES measures. It has also problems in validity of reporting.

Occupation reflects the privileges related to social standing, material resources, and job-related risk factors. Occupation-based indicators of SEP are widely used in the epidemiologic research due to their large availability in many routine data sources, including census data and death certificates. The individual current and the longest-held occupation are often taken into consideration to assess adult SEP. Measures from one or several individuals belonging to the same family unit can be used to characterize the SEP of others connected to them, e.g., children, spouse, elderly, unemployed. Among the limitations, occupation indicators clearly cannot be assigned to currently unemployed or retired people, housekeepers, students, and people with informal, unpaid, or illegal jobs. Also, classification for some job categories is difficult. Moreover, the definition of occupation related to SEP may have different meanings according to individual birth date and geographical location, which consequently represents an issue in terms of international comparisons.

Education reflects the intellectual assets of individuals besides the socioeconomic conditions in childhood and adolescence and it represents people potential opportunity to access, in the future, to higher level jobs and earnings. Educational attainment is a widely used indicator of SEP. The strength of using education as a proxy for SEP in the adult population is its smaller likelihood of reverse causation (e.g., whether poor health may be cause or consequence of low SEP), which always represents a big issue of other standard SEP measures. Indeed, it is generally assumed that the cycle of education is complete, or otherwise identifiable, before health issues may occur (Galobardes et al. 2006; Shavers 2007). In many epidemiological studies where measures of income, status, and occupation are not available, educational level is frequently used as the social position indicator and it tends to be empirically associated with the other measurements (d'Errico et al. 2017). The value of this social indicator, however, varies across geographic areas and cohorts.

3 Description of Two Worldwide Epidemiological Data Consortia

The Stomach Cancer Pooling (StoP) Project and the International Head and Neck Cancer Epidemiology (INHANCE) are an example of two large data consortia, in which the University of Milan is proactively involved in. In particular, our Department is the coordinator center of the StoP project and has been promoted several investigations and statistical data analyses based on the INHANCE consortium. Brief descriptions on these data consortia are here reported.

3.1 The StoP Project Consortium

The StoP Project is a consortium of epidemiological studies on gastric cancer established in 2012; the University of Milan is among the founders of the project. Up to date, the consortium includes 33 studies for a total of 12,753 gastric cancer cases and 30,682 controls. Of the patients, 40% are from Asia, 43% from Europe, and 17% from North America; 34% are women and 66% men; the median age is 61 years (Pelucchi et al. 2015).

The main aim of the StoP Project is to examine several lifestyles, including SEP, environmental, and genetic risk factors for gastric cancer, taking advantage of a large data set with original information from various geographic areas. The statistical analyses are carried out through pooled analyses of individual-level data, after central collection and validation of the original datasets. As compared to meta-analyses, the individual-level data approach allows harmonization of information and analyses, consistency of adjustment terms and multi-variate models, and investigation of heterogeneity and interaction between covariates (Ioannidis et al. 2013).

The StoP project challenge is therefore to improve knowledge of the etiology of gastric cancer, allowing decision-makers to plan preventive strategies, and providing a contribution to its control and its impact on the health of our population (Pelucchi et al. 2015; Winn et al. 2015).

3.1.1 Definition of SEP in StoP Project Consortium

The uniquely large sample size and the access to raw patient-level data allowed the StoP consortium to accurately assess the relation of SEP with gastric cancer overall and its subsites and histological subtypes, as well as to assess the associations in subgroups of the population according to sex, age, geographic area, and macroeconomic measure of income inequality of the country.

The level of education and household income were considered as proxies for the SEP. A uniform definition of occupational position among the included studies was not available at the time of the analysis, making unfeasible the evaluation of the relationship between occupational-based social class and gastric cancer risk. Education data were standardized across studies following the International Standard Classification of Education from the United Nations Educational, Scientific and Cultural Organization (UNESCO). This international reference classification facilitates comparisons between education systems across countries worldwide. Specifically, ISCED 2011 (UNESCO 2012) was issued as reference in the StoP project consortium. Education level was divided into three categories: (i) low education level, including no education, early childhood, and primary education (ISCED 0–1); (ii) intermediate education level, including secondary education (lower and upper) and postsecondary non-tertiary education (ISCED 2–4); (iii) high education level, including tertiary vocational and higher education, often designed to provide participants with professional knowledge, skills and competencies and education leading to a university degree (ISCED 5–6).

Household income was estimated by standardizing available study questionnaires data; comparable income levels were grouped into four categories, i.e., low, lower middle, upper middle, and high (Rota et al. 2019).

3.2 The INHANCE Consortium

The INHANCE consortium was established in 2004 as a collaboration among international research groups and includes investigators from over 35 international studies who have pooled their data on 30,000 patients with head and neck cancer and 40,000 controls without these cancers.

The primary goal of the consortium is to address the associations of head and neck cancer with a number of environmental factors, in particular tobacco smoking and alcohol drinking (i.e., the most relevant risk factors for the disease). The large sample size achieved by pooling studies allowed to assess the role of anthropometric characteristics, nutritional factors, income, and education. Moreover, INHANCE has the sufficient sample size to investigate subtypes of head and neck cancer (specifically oral cavity, oropharyngeal, hypopharyngeal, and laryngeal cancers) and to study heterogeneity in results across studies, geographic areas, and time periods, which may help to better identify unique risk factors or vulnerable populations (Winn et al. 2015).

3.2.1 Definition of SEP in INHANCE Consortium

On the basis of its information-rich data sets, the INHANCE consortium performed a detailed study with the aim to assess the risk for head and neck cancer associated with low educational status and household income. Analyses were carried on the overall database and by age, sex, cancer subsite, geographic location, and macroeconomic measure of income inequality at country-level.

Education data were standardized across studies following the International Standard Classification of Education from UNESCO. ISCED 97 protocol was used to categorize education levels (UNESCO 1997), which were divided into three strata: (i) low education level, including no education, early childhood, and primary education (ISCED 0–1); (ii) intermediate education level, including secondary education (lower and upper) and postsecondary non-tertiary education (ISCED 2–4); (iii) high education level, which comprised further education including vocational education and higher education including university degree (ISCED 5–6).

Concerning household income, in the INHANCE consortium, data were standardized as far as possible (i.e., when in the original study questionnaire the proper categorization was addressed) by grouping comparable levels based on the strata used in the original study, starting from category 1 associated to the lowest income levels, up to the highest within category 5 (Conway et al. 2015).

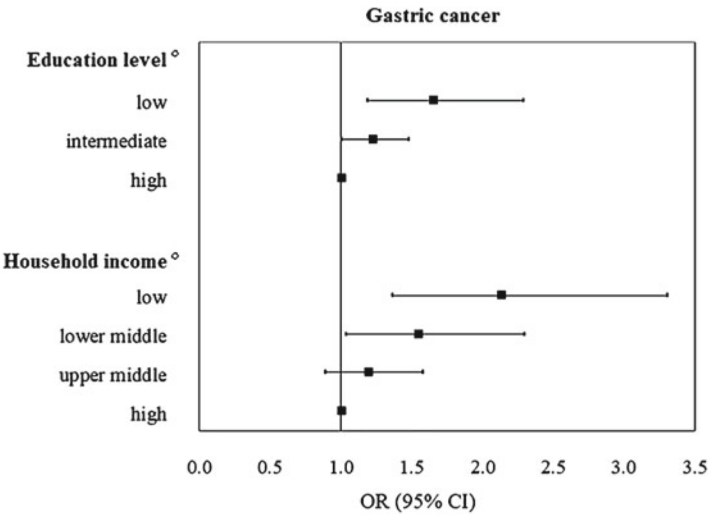
4 Results

4.1 Results from the StoP Project Consortium

Findings from the StoP consortium showed that SEP, measured through education level and household income, is a strong determinant of gastric cancer.

Data on education level were available from 25 out of 33 studies participating in the StoP consortium (11 from European countries, 6 from Asian countries, 3 from North, and 5 from Central/South American countries), for a total of 10,000 gastric cancer cases and 25,000 healthy controls. Seven studies (4 from Asian countries, 2 from Brazil, and 1 from Canada) provided data on household income.

To analyze the association of education and household income with gastric cancer risk, we firstly estimated study-specific Odds Ratios (OR) and the corresponding 95% Confidence Intervals (CI) using multi-variable unconditional logistic regression models. Analyses showed that subjects with intermediate and low education levels had, respectively, 22% (pooled OR, 1.22, 95% CI, 1.01–1.48) and 65% (pooled OR 1.65, 95% CI, 1.19–2.29) increased risks of gastric cancer compared to those with higher education attainment (Fig. 1). Results were adjusted for a number of lifestyle and dietary habits, which may confound the associations of SEP with gastric cancer, including tobacco smoking, race/ethnicity, and the intake of alcohol, fruit, and vegetables. Strong positive associations were observed for both cardia and non-cardia gastric cancers, as well as for diffuse and intestinal subtypes. In addition, the positive association between education level and gastric cancer risk was evident regardless of infection with *Helicobacter Pylori* (*HP*), and in subgroups defined by age, sex, cigarette smoking, and alcohol drinking. In analyses by geographic area, strong positive associations were reported by studies from Europe and Asia, while combined results from the three North American studies indicated a non-significant positive association. Conversely, Central/South America studies (mainly Mexican



Note: ° Adjusted for age, sex, study center, alcohol drinking, tobacco smoking, race/ethnicity, fruit and vegetable consumption. Education was standardized using the International Standard Classification of Education (ISCED 2011). Low education corresponds to ISCED 0–1, Intermediate education to ISCED 2–4 and High education to ISCED 5–6.

Fig. 1 Pooled odds ratios (OR) and 95% confidence intervals (CI) of stomach cancer (Stomach cancer Pooling [StoP] Project consortium) according to education level and household income. The reference category is high level

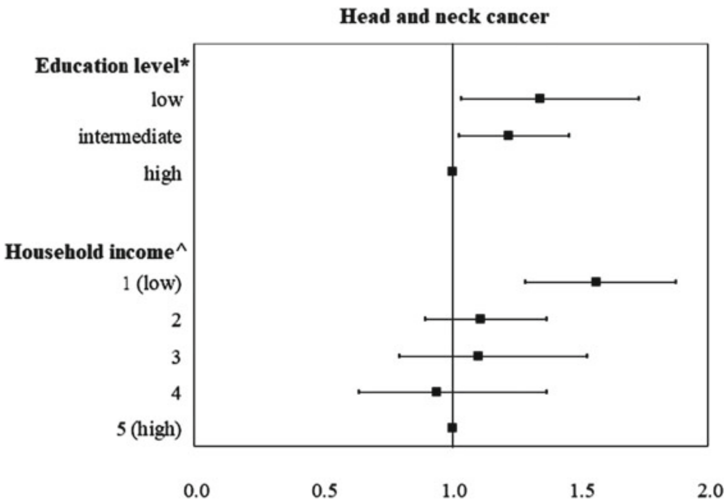
studies) did not find any relation between education level and gastric cancer, raising concerns about the reliability of education as a proxy for the SEP in such countries. Mexico has high rates of income inequality and wealth is concentrated in a small fraction of the population, while the majority is poor and has limited access to education, and thus better living conditions. Large segments of the population still fail to achieve even basic education. Alternatively, education may be a better indicator of SEP in high- than in middle-income countries. In low- and middle-income countries, in fact, education is strongly related to social class in childhood, while in high-income countries it mainly reflects physical or psychological impairments that in the long term may influence cancer risk to a greater extent. When household income was used as proxy of SEP, a 35% increased risk of gastric cancer was observed for subjects in the lowest versus the highest income category (OR 2.13, 95% CI, 1.37–3.31) (Fig. 1) (Rota et al. 2019).

4.2 Results from the INHANCE Consortium

The INHANCE consortium indicated that low education level and low income are risk factors for head and neck cancer, even in the absence of the well-known lifestyle risk factor for this cancer, namely smoking, use of other tobacco products and alcohol drinking.

The estimated study-specific OR and 95% CI for the association of education and income for head and neck cancer were calculated using unconditional logistic regression based on 31 case–control studies and almost 24,000 head and neck cancer patients and 32,000 controls. The analyses indicated that subjects with low education had a more than two-fold increased risk of head and neck cancer compared to those with high education (pooled OR 2.50), after allowance for age and sex. The risk for subjects in the intermediate education category was increased by 80%. When accounting for smoking, alcohol, and selected dietary factors, the association was attenuated but still significant, with an over 30% elevated risk among subjects with low versus those with high education (pooled OR 1.34, 95% CI, 1.04–1.73) (Fig. 2). In addition, the risk remained increased by over 50% in subjects who never smoked or used other type of tobacco and never drank alcohol (OR 1.61, 95% CI, 1.13–2.31). This suggests that the association of head and neck cancer with education level is not totally attributable to these detrimental behaviors, although some degree of residual confounding could not be excluded. In addition, part of the association observed with education could be explained by *Human Papilloma Virus* (HPV) infection.

The association with low education level was observed for all head and neck cancer subsites (i.e., oral cavity, oropharynx, hypopharynx, and larynx), and was



Note: * Adjusted for age, sex, study center, alcohol drinking, tobacco, fruit and vegetabl e consumption. Education was standardized using the International Standard Classification of Education (ISCED 2011). Low education corresponds to ISCED 0–1, Intermediate education to ISCED 2–4 and High education to ISCED 5–6.
^ Adjusted for age, sex, study center, alcohol drinking and smoking.

Fig. 2 Pooled odds ratios (OR) and 95% confidence intervals (CI) of head and neck cancer (International Head and Neck Cancer [INHANCE] consortium) according to education level and household income. The reference category is high level somehow stronger in North American and Central/South American populations as well as in higher income inequality countries.

The analyses on household income were based on 10 studies with available information (9 from USA and 1 from Porto Rico). Results were in line with those for education level, with an over two-fold increased risk for the lowest *vs* the highest category of income, in an analysis which takes into account age and sex. Again, the association was attenuated, but still evident, after allowance for smoking and alcohol, with an over 50% increased risk among subjects with the lower monthly income (Fig. 2) (Conway et al. 2015).

5 Conclusion

Social epidemiology is crucial to understand the sociostructural factors related to health and disease. In an era of fast inter-diffuse communication and data-sharing, large collaborative groups and data consortia are among the most effective strategies to create new social epidemiological useful evidences. In particular, data analyses of large epidemiological consortia found that SEP is strongly related to a number of cancers. Notably, the results from two large epidemiological consortia on gastric (StoP) and head and neck cancers (INHANCE) indicated that the association with SEP persists even after allowance for smoking and alcohol, i.e., unfavorable correlates of most cancer types and associated to SEP, suggesting that the SEP-cancer association follows pathways beyond these detrimental behaviors. Up to date, most industrialized countries will be challenged to identify such influencing factors and their means of operating throughout the whole of society. Reduction of socioeconomic inequalities both at national and international level is advocated to decrease the burden of cancers in deprived populations.

References

Bennett, J. E., Stevens, G. A., Mathers, C. D., Bonita, R., Rehm, J., Kruk, M. E., et al. (2018). NCD countdown 2030: Worldwide trends in non-communicable disease mortality and progress towards Sustainable Development Goal target 3.4. *Lancet*. https://doi.org/10.1016/s0140-6736(18)31992-5.

Conway, D. I., Brenner, D. R., McMahon, A. D., Macpherson, L. M. D., Agudo, A., Ahrens, W., et al. (2015). Estimating and explaining the effect of education and income on head and neck cancer risk: INHANCE consortium pooled analysis of 31 case-control studies from 27 countries. *International Journal of Cancer*. https://doi.org/10.1002/ijc.29063

d'Errico, A., Ricceri, F., Stringhini, S., Carmeli, C., Kivimaki, M., Bartley, M., et al. (2017). Socioeconomic indicators in epidemiologic research: A practical example from the LIFEPATH study. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0178071

Galobardes, B., Shaw, M., Lawlor, D. A., Lynch, J. W., & Davey Smith, G. (2006). Indicators of socioeconomic position (part 1). *Journal of Epidemiology and Community Health.* https://doi.org/10.1136/jech.2004.023531

Geyer, S., Hemstrom, O., Peter, R., & Vagero, D. (2006). Education, income, and occupational class cannot be used interchangeably in social epidemiology. Empirical evidence against a common practice. *Journal Epidemiol Community Health.* doi:10.1136/jech.2005.041319.

Honjo, K. (2004). Social epidemiology: Definition, history, and research examples. *Environmental Health and Preventive Medicine.* https://doi.org/10.1007/BF02898100

Ioannidis, J. P., Schully, S. D., Lam, T. K., & Khoury, M. J. (2013). Knowledge integration in cancer: Current landscape and future prospects. *Cancer Epidemiology, Biomarkers & Prevention.* https://doi.org/10.1158/1055-9965.EPI-12-1144

Marmot, M. A., & Allen, J. J. (2014). Social determinants of health equity. *American Journal of Public Health.* https://doi.org/10.2105/AJPH.2014.302200

Niessen, L. W., Mohan, D., Akuoku, J. K., Mirelman, A. J., Ahmed, S., Koehlmoos, T. P., et al. (2018). Tackling socioeconomic inequalities and non-communicable diseases in low-income and middle-income countries under the sustainable development agenda. *Lancet*. https://doi.org/10.1016/S0140-6736(18)30482-3

NIH. (2019). Consortia to advance collaboration in epidemiologic and cancer research. Retrieved July 22, 2019, from https://epi.grants.cancer.gov/Consortia/.

Pelucchi, C., Lunet, N., Boccia, S., Zhang, Z. F., Praud, D., Boffetta, P., et al. (2015). The stomach cancer pooling (StoP) project: Study design and presentation. *European Journal of Cancer Prevention.* https://doi.org/10.1097/CEJ.0000000000000017

Rota, M., Alicandro, G., Pelucchi, C., Bonzi, R., Bertuccio, P., Hu, J., et al. (2019). Education and gastric cancer risk-an individual participant data meta-analysis in the StoP project consortium. *Internatonal Journal of Cancer*. https://doi.org/10.1002/ijc.32298

Shavers, V. L. (2007). Measurement of socioeconomic status in health disparities research. *Journal of the National Medical Association, 99*(9), 1013–1023.

UNESCO. (1997). *International Standard Classification of Education: ISCED 1997*. Paris: UNESCO Institute for Statistics.

UNESCO. (2012). *International standard classification of education: ISCED 2011*. Montreal: UNESCO Institute for Statistics.

Vaccarella, S., Lortet-Tieulent, J., Saracci, R., Conway, D. I., Straif, K., & Wild, C. P. (2019). *Reducing social inequalities in cancer evidence and priorities for research.* Geneve: IARC Scientific Publication.

Wagstaff, A. (2002). Poverty and health sector inequalities. *Bulletin of the World Health Organization, 80,* 97–105.

WHO. (2018a). Cancer—Key facts. Retrieved July 22, 2019, from https://www.who.int/news-room/fact-sheets/detail/cancer.

WHO. (2018b). Noncommunicable diseases. Retrieved July, 22, 2019, from https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases.

Winn, D. M., Lee, Y. C., Hashibe, M., & Boffetta, P. (2015). The INHANCE consortium: Toward a better understanding of the causes and mechanisms of head and neck cancer. *Oral Diseases.* https://doi.org/10.1111/odi.12342