STUDY PROTOCOL

Unveiling the alcohol harm paradox: analysis of socioeconomic differences in alcohol harm in Finland in a nationwide cohort study

Sponsored by:
Finnish Instituted for Health and Welfare

Collaborators:
The Social Insurance Institution of Finland (KELA)

Funded by:
The study is funded by the Finnish Sustainable Growth Programme (RRP) from the Ministry of Social Affairs and Health and the National Health Index project funded by the Social Insurance Institution (grant 72/26/2021).

Authors
Sebastián Peña,[1] Elsi Lindell,[1] Sonja Lumme,[1] Laura Kestilä,[1] Kati Sarnola,[2] Marke Jääskeläinen,[1] Päivikki Koponen[1]

1. Department of Health and Welfare, Finnish Institute for Health and Welfare, Mannerheimintie 166, Helsinki, Finland
2. The Social Insurance Institution of Finland (KELA), Nordenskiöldinkatu 12, Helsinki, Finland

FINAL version 1.0
This protocol has not been revised
Date of document: April 25, 2023

Background

Alcohol consumption is a leading risk factor for death and disability,[1] resulting in substantial societal costs[2] and social and community harm.[3] Alcohol-related harm has been consistently shown to be higher among people of lower socioeconomic status, despite the fact they often report similar or lower levels of alcohol use.[4] This apparent contrast has been called the alcohol harm paradox.

Understanding what explains the alcohol harm paradox is crucial for the design and implementation of population health interventions to reduce socioeconomic differences in alcohol harm. Explanations for the paradox can be broadly categorised into three groups: (i) differential exposure to alcohol,[5] drinking patterns[6,7] and trajectories and to other behavioural risk factors and joint effects among them,[6,8,9] (ii) differential vulnerability resulting from individual factors, such as biological characteristics, psychological traits or stress, cumulative disadvantage[10] and broader community and societal upstream factors,[11] and (iii) differential biases in the measurement of alcohol exposure.[12]

A crucial weakness of existing empirical studies comes from the operationalization of alcohol harm. With few exceptions,[13] a vast majority of studies used a composite endpoint combining several causes of alcohol-attributable deaths or hospitalizations or even merging alcohol-attributable hospital admissions with deaths in a single outcome.[6,9] While this strategy increases the statistical power to analyse rare events, composite endpoints are prone to misclassification bias by masking divergent underlying patterns and associations.[14]

Cause-specific analyses might shed light on different mechanisms driving the socioeconomic differences in overall alcohol-attributable harm, as well as opening potential avenues for policy interventions to reduce them. We will take advantage of a recently formed dataset covering the total Finnish population to explore associations between socioeconomic status and cause-specific alcohol-attributable events. An additional contribution to the literature will be reporting both relative and absolute inequalities in alcohol-attributable harm. Absolute differences have been rarely reported[8,15] even though might be as relevant as relative differences because, from a pragmatic standpoint, they are more feasible to reduce.[16]

The study will aim to (1) examine the relative and absolute socioeconomic differences in cause-specific alcohol-attributable hospital admissions and deaths; (2) describe the geographical differences in cause-specific alcohol-attributable hospital admissions and deaths; and (3) quantify the relative contribution of each specific alcohol-attributable cause to the overall alcohol-attributable harm.

Methods

Study design and setting

The study is a nationwide population cohort study of the entire Finnish population from January 1, 2016, to 28 February 2020. Individuals in Finland are assigned a unique identifier at birth or immigration and used in all public and private administrative records. We will use the unique identifier to link socioeconomic information with hospital admissions and deaths.

Participants

Participants will be all permanent residents (see below for a definition) aged 25 and over in Finland by the last day of 2015 and without an alcohol-attributable hospital admission during the past three years (January 2013-December 2015). The age limit of 25 was considered because younger adults might have their education ongoing and their income and occupational class may not reflect their true socioeconomic standing. We will exclude residents with prior hospital admissions during the past three years to reduce the risk of reverse causality.

Data sources

We will use data from the FOLK database (basic module, module for Household Dwelling Unit and module of Degree/Qualifications) maintained by Statistic Finland.[17] The FOLK database contains individual-level data on income, education, family, degrees and qualifications and cohabitation. Educational level data comes from the Register of Completed Education and Degrees. Income data derive from Statistics Finland's income distribution statistics and personal taxation data. Occupation data derives from a combination of sources from registers of employment relationships, statistics on wages and salaries and data on enterprises.[18] We will use data on hospital admissions from the Finnish Care Register for Health Care (HILMO) which covers all hospital admissions in public and private inpatient hospitals in Finland. We will use mortality data from the Causes of Death register administered by Statistics Finland.

Outcomes

The outcome will be cause-specific alcohol-attributable hospital admissions and deaths. We will include any wholly-attributable cause (i.e. an attributable fraction of 1) included in the main and secondary causes of hospitalization as well as the underlying and contributory causes of death. Drawing on national reports and previous studies,[19] we will use the following International Classification of Disease (ICD) codes:

ICD-10 code	Diagnosis
F10.0-1	Alcohol intoxication
F10.2	Alcohol dependence
F10.3-9	Psycho-organic syndrome caused by alcohol
K70	Alcoholic liver disease
K85.2, K86.0	Pancreatic diseases caused by alcohol
K29.2	Gastritis caused by alcohol
I42.6	Alcoholic cardiomyopathy
G312, G4051, G621, G721	Degeneration of the nervous system due to alcohol, Epileptic seizures related to alcohol, Alcoholic polyneuropathy, Alcoholic myopathy
T51, X45	Alcohol poisoning

We will exclude perinatal causes (O354, P043, Q860) as they are impossible in this study population.

Hospital admissions. We will define hospital admission as any event including at least one overnight stay in inpatient care. Patients having two hospital admissions in two hospitals with overlapping dates will be considered as a single hospital admission. We will include admissions where the main or secondary causes are attributable to alcohol. Given we are examining cause-specific alcohol-attributable events, we defined criteria to attribute the cause in case of overlapping or multiple alcohol-attributable ICD-10 codes. For hospital admissions with more than one alcohol-attributable cause, we will consider the main cause as the hospitalization cause. In hospital admissions with more than one alcohol-attributable ICD-10 code, we will consider the main cause as the hospitalization cause. If there are several alcohol-attributable main causes, we will consider first the complexity of the inpatient provider (provider code in the HILMO dataset). In the case of ties by complexity, we will consider the longest stay. We will not include outpatient visits, day-hospital treatments or emergency care visits.

Deaths. We will define alcohol-attributable deaths as any death where an alcohol-attributable cause is either the underlying, immediate, intermediate or contributory cause of death.

We will follow participants until the first hospital admission, death or end of follow-up. We will restrict follow-up until February 28, 2020, to avoid confounding due to the COVID-19 pandemic.

Exposure

The exposure of interest is socioeconomic status. Our main proxy of socioeconomic status will be disposable household income. In addition, we will use educational level and occupational class. All exposures are measured by the last day of the calendar year (2015). We will calculate equivalised disposable income by adjusting the combined disposable income of a household to the household size unit using the OECD-modified equivalence scale. The income measure in this study is the sum of wage and salary earnings, self-employment income, property income and current State transfers after subtracting direct taxes and social security contributions. For persons in non-dwelling units (e.g. living in institutions, homeless, abroad or registered as unknown) we will consider their personal net income without equivalization. We will also run sensitivity analyses excluding those participants.[20]

We will categorise educational level according to the highest qualification completed degree into lower secondary education or less (missing data on education), upper secondary education or post-secondary (levels 3 and 4) and or short-cycle tertiary education, Bachelor's or equivalent level and higher (levels 5, 6, 7 and 8). Due to restrictions in available data, we do not have information on individuals' educational attainment if it is less than lower secondary education. Because of this, we do not know the proportion of the study population the information is truly missing.

We will categorise occupational class into manual workers (codes 51, 52, 53, 54, 59), self-employed persons (codes 10 and 20), lower-level employees (codes 41, 42, 43 and 44) and upper-level employees (31, 32, 33, 34), students, pensioners and those unemployed. This classification is created by Statistics Finland and derives from the person's main type of activity, occupation, occupational status and industry. Persons outside the labour force are assigned the same class as the reference person in the household. For persons in non-dwelling units, given there is no reference person, we will treat this group as a separate category of occupational class. The Supplementary Appendix includes a detailed description of how these variables are created.

Confounders

We will control for sex, age, marital status, and mother tongue. We will categorise sex as male or female. We will categorise marital status as unmarried, married, in a registered partnership or separated, divorced or divorced from a registered partnership or widowed or widowed from a registered partnership. We will categorise mother tongue as Finnish, Swedish and other.

Bias

Confounding bias. We will reduce the risk of confounding bias by adjusting for potential confounders that are associated with the exposure and the outcome. Our data derives from national registries and provides excellent coverage of observed sociodemographic characteristics. However, we do not have data on behavioural factors, such as tobacco or

alcohol use, or personality traits. The risk of residual confounding can only be attenuated but not be fully removed in our study design.

Information bias. Information from national registries is not subject to recall or social desirability bias. Education data can be subject to measurement error due to missing information for lower levels of education. This can be more severe for immigrants because the Register of Education and Completed Degrees might not register degrees obtained abroad. Statistics Finland has been able to obtain information on more than 13.000 degrees obtained in Nordic countries and 26.000 degrees in other countries.[21] Nonetheless, the educational attainment of foreigners who studied abroad might be underestimated. In sensitivity analyses, we will re-run the analyses excluding foreigners. We anticipate, however, that this will have a negligible effect on our estimates, as foreigners represent a small proportion of the population. The Finnish economy is highly formalized and informal employment is estimated at 1-3% of those employed. These include freelancers, own-account workers in the informal economy, contributing family workers and grant recipients. Therefore, income and employment statistics include 97-99% of the employed population.[22] Statistics Finland also has routine procedures to assess the quality of registry data, starting from the studies carried out to compare census with survey data in the 1980s and from employment surveys conducted regularly.[18] The quality of Finnish registry data is, hence, subject to minimal information bias.

Selection bias. Our data comes from the total Finnish population and, therefore, is not subject to selection due to participation. The data includes permanent residents in Finland, defined as all Finnish nationals and foreign nationals who have a legal domicile in Finland and intend to stay (or have stayed) for at least one year. This excludes foreign nationals living in Finland for less than a year, asylum seekers who have not been granted a legal domicile and temporary migrant workers (e.g. cherry pickers and construction workers), and Finnish nationals living temporarily abroad.[23]

Selection bias can also arise due to loss of follow-up, although in Finland deaths missing a death certificate are extremely rare (about 0.2%).[24] We consider the risk of selection bias due to loss of follow-up relatively small.

Quantitative variables

All variables included in the analyses are categorical.

Ethical approval

The use of administrative records does not require ethical approval in Finland. All data used was pseudonymised and linked by Statistics Finland and will be analysed in a protected remote access system FIONA. We will comply with Finnish regulations established under the National Statistics Act and the supplementary National Data Protection Act as well as with the General Data Protection Regulation from the European Union.[25] Permission for data usage and linkage was given to the Finnish Institute for Health

and Welfare as part of the Compensation Criteria and Evaluation of Provinces project (permission K/3184/07.03.00/2022, U1098_al9)

Statistical methods

We will use the relative index of inequality (RII) and the slope index of inequality (SII) to compare and quantify the socioeconomic gradient in cause-specific alcohol-attributable hospital admissions and deaths. We will use regression models for estimation with individual time-to-event data: Cox proportional hazards models for relative inequalities and Aalen additive hazard models for absolute inequalities.

For the RII, we will fit the following Cox proportional hazards model,

(1) $$\lambda(t\,|G = k) = \beta_0(t)exp\{\beta x_{(k)} + \beta(t)L_i\}$$

Where G represents the socioeconomic group and we approximate the rank of each person in the kth is socioeconomic group by the rank $x_{(k)}$ defined as the percentage of the population in strictly higher groups plus half of the percentage in group k, and L_i denotes a vector of covariates (sex, marital status, etc.).[26] This method allows us to incorporate the whole socioeconomic gradient, instead of comparing the lowest and highest socioeconomic groups. The estimate of $\widehat{RII}$ is equivalent to $exp(\widehat{\beta})$.[26]

We will use age as the timescale, as age is strongly correlated with the risk of hospitalization and mortality.[27] We will report regression estimates as hazard ratios (HRs) with 95% confidence intervals (CIs). Given we studied cause-specific events, deaths due to other causes represent competing risks. To account for this, we will use cause-specific hazards by censoring other-cause events to the time of the death by other causes or the end of follow-up on February 28, 2020. We will examine the proportional hazards assumption globally and for each covariate using scaled Schoenfeld residuals.[28]

For the SII, we will fit age-standardized Aalen additive hazard models,

(2) $$\lambda(t\,|G = k) = \alpha_{0s}(t) + \alpha_s x_{(k)} + \alpha(t)L_i$$

where the age standardization is the weighted sum of estimated age-group-specific indices. For this we will fit separate additive hazard models within each age-group $s = 1,\ldots,S,$. The estimate of the $\widehat{SII}$ is equal to $\widehat{\alpha}$.

Same as before, we will use age as the timescale and report regression estimates as events per 100,000 person-years with 95% confidence intervals. We will report cause-specific hazards to account for competing risks.

In practice, we will run a model for each cause-specific condition (nine models in total) separately for alcohol-attributable hospital admissions and alcohol-attributable deaths.

For objective 2, we will run stratified analyses by hospital districts and Åland (n =21). We will report the RII and SII by region in plots or maps in order to obtain a visual representation of both the degree of between-region inequalities and geographical patterns in the distribution of alcohol-attributable harm. Considering we might not have enough statistical precision to obtain informative estimates for regional differences for all cause-specific outcomes, we will report overall alcohol-attributable admissions and deaths (separately) and, in the Supplementary Appendix, the most common specific causes.

For objective 3, we will quantify the relative contribution of each specific cause by dividing the SII for the specific cause by the overall SII for either hospital admission or death and multiplying by 100.

We will run the same Cox proportional hazards and additive hazards models for educational level and occupational class. However, given these are categorical variables with an unclear ordinality and the relationship with the outcome might not be linear, we are not able to estimate the RII or SII. Instead, we will use the standard approach of comparing the hazard ratio and additive hazard comparing the category to a reference group (highest education and upper-level employees, respectively).

Missing data for covariates in the FOLK datasets is minimal. Sex, age, marital status and mother tongue do not have missing data. We described above how we will handle missing data for education. For models using income and occupation, we will use multiple imputation using chained equations to address missing data.

We will carry out additional analyses separating hospital admissions and deaths where alcohol is the main cause of hospitalization and the underlying cause of death versus a secondary or contributory cause of death. Alcohol use is more clearly a sufficient component cause in the main and underlying causes of admissions and death, and we would expect stronger socioeconomic differences.

We will use SAS version 9.4 and R version 3.6.2 for all analyses.

References

1. Griswold, M. G. *et al.* Alcohol use and burden for 195 countries and territories, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *The Lancet* **392**, 1015–1035 (2018).
2. Manthey, J. *et al.* What are the Economic Costs to Society Attributable to Alcohol Use? A Systematic Review and Modelling Study. *PharmacoEconomics* **39**, 809–822 (2021).
3. Room, R. *et al.* The Drinker's Effect on the Social Environment: A Conceptual Framework for Studying Alcohol's Harm to Others. *Int. J. Environ. Res. Public. Health* **7**, 1855–1871 (2010).
4. Probst, C., Lange, S., Kilian, C., Saul, C. & Rehm, J. The dose-response relationship between socioeconomic deprivation and alcohol-attributable mortality risk-a systematic review and meta-analysis. *BMC Med.* **19**, 268 (2021).
5. Probst, C., Kilian, C., Sanchez, S., Lange, S. & Rehm, J. The role of alcohol use and drinking patterns in socioeconomic inequalities in mortality: a systematic review. *Lancet Public Health* **5**, e324–e332 (2020).
6. Sydén, L., Sidorchuk, A., Mäkelä, P. & Landberg, J. The contribution of alcohol use and other behavioural, material and social factors to socio-economic differences in alcohol-related disorders in a Swedish cohort. *Addict. Abingdon Engl.* **112**, 1920–1930 (2017).
7. Mäkelä, P. & Paljärvi, T. Do consequences of a given pattern of drinking vary by socioeconomic status? A mortality and hospitalisation follow-up for alcohol-related causes of the Finnish Drinking Habits Surveys. *J. Epidemiol. Community Health* **62**, 728–733 (2008).
8. Peña, S. *et al.* Joint effects of alcohol use, smoking and body mass index as an explanation for the alcohol harm paradox: causal mediation analysis of eight cohort studies. *Addict. Abingdon Engl.* **116**, 2220–2230 (2021).
9. Katikireddi, S. V., Whitley, E., Lewsey, J., Gray, L. & Leyland, A. H. Socioeconomic status as an effect modifier of alcohol consumption and harm: analysis of linked cohort data. *Lancet Public Health* **2**, e267–e276 (2017).
10. Mäkelä, P. Alcohol-related mortality as a function of socio-economic status. *Addict. Abingdon Engl.* **94**, 867–886 (1999).
11. Boyd, J. *et al.* Causal mechanisms proposed for the alcohol harm paradox-a systematic review. *Addict. Abingdon Engl.* **117**, 33–56 (2022).
12. Peña, S. *et al.* Measurement error as an explanation for the alcohol harm paradox: analysis of eight cohort studies. *Int. J. Epidemiol.* **49**, 1836–1846 (2021).
13. Sadler, S. *et al.* Understanding the alcohol harm paradox: an analysis of sex- and condition-specific hospital admissions by socio-economic group for alcohol-associated conditions in England. *Addict. Abingdon Engl.* **112**, 808–817 (2017).
14. Goldberg, R., Gore, J. M., Barton, B. & Gurwitz, J. Individual and Composite Study Endpoints: Separating the Wheat from the Chaff. *Am. J. Med.* **127**, 379–384 (2014).
15. Mackenbach, J. P. *et al.* Inequalities in Alcohol-Related Mortality in 17 European Countries: A Retrospective Analysis of Mortality Registers. *PLoS Med.* **12**, e1001909 (2015).

16. Mackenbach, J. P. Should we aim to reduce relative or absolute inequalities in mortality? *Eur. J. Public Health* **25**, 185–185 (2015).
17. Statistics Finland. *FOLK Basic data - data description*.
18. Statistics Finland. Employment: documentation of statistics. https://tilastokeskus.fi/en/statistics/documentation/tyokay.
19. Koskinen, H. & Virtanen, S. *Yearbook of Alcohol and Drug Statistics 2022*. (Finnish Institute for Health and Welfare, 2022).
20. Suokas, K. *et al.* Association of Income With the Incidence Rates of First Psychiatric Hospital Admissions in Finland, 1996-2014. *JAMA Psychiatry* **77**, 274–284 (2020).
21. Statistics Finland. Educational structure of population: documentation of statistics. https://tilastokeskus.fi/en/statistics/documentation/vkour.
22. Sutela, H. Informal work in Finland.
23. Statistics Finland. Concepts and definitions. https://www.stat.fi/tup/maahanmuutto/kasitteet-ja-maaritelmat_en.html.
24. Statistics Finland. Quality Description: Causes of death 2020. https://www.stat.fi/til/ksyyt/2020/ksyyt_2020_2021-12-10_laa_001_en.html.
25. Kankaanranta, P. & Melakari, A. Creating ready-made research datasets from national administrative registers. in *Expert Meeting on Statistical Data Confidentiality* (2021).
26. Moreno-Betancur, M., Latouche, A., Menvielle, G., Kunst, A. E. & Rey, G. Relative Index of Inequality and Slope Index of Inequality: A Structured Regression Framework for Estimation. *Epidemiology* **26**, (2015).
27. Austin, P. C., Lee, D. S. & Fine, J. P. Introduction to the Analysis of Survival Data in the Presence of Competing Risks. *Circulation* **133**, 601–609 (2016).
28. Grambsch, P. & Therneau, T. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika* **81**, 515–526 (1994).

Supplementary Appendix to the protocol

Unveiling the alcohol harm paradox: analysis of socioeconomic differences in alcohol harm in Finland in a nationwide cohort study

Sebastián Peña, Elsi Lindell, Sonja Lumme, Laura Kestilä, Kati Sarnola, Marke Jääskeläinen, Päivikki Koponen

Each of the three exposures are defined using the study population's status at the baseline. Educational level and occupational status are defined as of 31.12.2015.

Income

Database source: FOLK module for household dwelling unit
Original variable names: Disposable money income (household-dwelling unit), Consumption units according to modified OECD-scale, Disposable income
Technical names: kturaha_ak, modeocd, kturaha_k

Statistics Finland maintains individual-level information on income in its Income distribution Statistics. As a measure of income level in this study, we use equivalised disposable income calculated for each individual belonging to a household dwelling unit. It is obtained by calculating the combined disposable income of a household and adjusting it by the number of consumption units in the household. We use the modified OECD scale of consumption units. Negative income is set to zero. The income describing variables (kturaha_ak and kturaha_k) are defined and composed the same way by Statistics Finland at the personal and household level.

For the analysis models, we categorise the individuals' income into deciles. The income decile limits are computed taking into account the total population aged 25 or older at the end of 2015. The equivalised disposable income is not available for those that do not belong to a household dwelling unit. These include e.g. institutionalised people, homeless and people registered as unknown. We will use the personal net income of these people without equivalisation.

Educational level

Database source: FOLK module for degree/qualification
Original variable name: Educational level of highest qualification/degree
Technical name: kaste

Information on individuals' highest completed degree or qualification is maintained by the National Register of Completed Education and Degrees. We use the definitions of Statistics Finland and recategorize the individuals' educational background into three categories as follows.

Category	Codes
Lower than upper secondary education	91 Level of education unknown Missing: no information on kaste can be linked
Upper secondary education	31 General upper secondary education 32 Vocational upper secondary education and training 33 Further qualification 41 Specialist qualification
Tertiary education or higher	51 Post-secondary non-higher vocational education 52 Lowest level tertiary education 61 Professional tertiary education 62 Polytechnic bachelor's degree 63 University bachelor degree 71 Higher polytechnic degree (Master) 72 Higher university degree (Master) 73 Professional specialisation in medicine, veterinary and dentistry 81 Licentiate's degree 82 Doctoral or equivalent level

Occupational status

Database source: FOLK module for basic data
Original variable name: Socioeconomic group
Technical name: sose

Statistics Finland formes occupational class using information on individual's economic activity, occupation and industry. We categorize occupational class according to Statistics Finland's definitions as follows.

Category	Codes
Self-employed persons	10 Farmers and forestry entrepreneurs 20 Self-employed persons, not in agriculture and forestry
Upper-level employees	31 Senior officials and upper management 32 Senior officials and employees in research and planning 33 Senior officials and employees in education and training 34 Other senior officials and employees
Lower-level employees	41 Supervisors 42 Clerical and sales workers, independent work 43 Clerical and sales workers, routine work 44 Other lower-level employees
Manual workers	51 Workers in agriculture, forestry and commercial fishing 52 Manufacturing workers 53 Other production workers 54 Distribution and service workers 59 Other workers
Students	60 Students
Pensioners	70 Pensioners
Unemployed	81 Unemployed

Others or unknown	82 Others (conscripts) 99 Unknown