Study protocol

Tobacco use and incidence of SARS-CoV-2 infection in the Finnish general population

Version 1.0, June 2, 2021

Revision history: this protocol has not been revised

Sebastián Peña,[1] Katja Ilmarinen,[1] Laura Kestilä,[1] Sakari Karvonen[1]

1. Department of Public Health and Welfare, Finnish Institute for Health and Welfare, Mannerheimintie 166, 00271, Helsinki, Finland

Collaborators:

Karolinska Institutet

Norwegian Institute of Public Health

Funding:

The Tobrisk-CoV study is funded by NordForsk (project number 105544)

Background

The COVID-19 pandemic is spreading rapidly worldwide. By June 2, 2021, there were more than 171 million confirmed SARS-COV-2 cases and more than 3.5 million deaths attributable to COVID-19 [1]. Researchers have explored potential risk factors to identify patients at high risk of infection or death, as well as for targeting pharmaceutical and preventive interventions [2].

Tobacco use, as a leading risk factor of death and disability due to respiratory diseases, was expected to increase the risk of SARS-CoV-2 infection and COVID-19 deaths [3, 4]. Smokers have generally increased risk of other respiratory infections and could be expected to have higher risk of SARS-CoV-2 due to repetitive hand-to-mouth handlings, increased mask handlings [5, 6], sharing of cigarettes and vape device [7] and creation of aerosols which might be carriers of viruses. On the other hand, smokers might have lower social contacts [8] and be less exposed to indoor places [9]. Studies have also reported mixed findings on whether tobacco or nicotine could modify the expression of ACE2 receptors, which provide a cellular entry point for SARS-CoV-2 [10, 11].

Earlier epidemiological studies showed that smokers were underrepresented among patients hospitalized due to COVID-19 [12, 13]. These surprising results could be explained by the selected nature of the sample or information bias arising from data collected retrospectively or from electronic health records [14, 15]. However, the most recent meta-analysis, including more diverse samples and study designs, has confirmed these early findings showing that current smokers had lower risk of SARS-CoV-2 infection than never smokers (Relative risk 0.71, 95% Credible interval 0.61; 0.82) [16].

A message of a protective effect of tobacco use could undermine public health efforts to curb its use and reduce the perception of harm in the general population [17]. Studies with general population samples and prospective data collection are thus urgently needed.

The aim of the study is to examine the association between tobacco use and the risk of SARS-CoV-2 infection. We will explore several forms of tobacco use (smoking, moist smokeless tobacco and e-cigarettes) and investigate whether introducing a potential collider bias by adjusting for mediating risk factors (alcohol use, physical activity and obesity) could have explained earlier results. We will use data from a prospective cohort study of nationally representative health surveys in Finland linked to SARS-CoV-2 incidence data, which is less subject to recall bias than previous case-control studies.

Methods

We will report the study in accordance with the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement for cohort studies [18].

Setting and study design

The design is a cohort study of cross-sectional population health surveys in Finland. The study populations were permanent residents in Finland from the FinSote surveys (2018, 2019 and 2020). We will link survey data to SARS-CoV-2 incidence data from the Communicable Diseases Registry until April 30, 2021, using a unique personal identifier assigned to all Finnish residents.

Data sources

We will use data from three cross-sectional population health surveys in Finland. The FinSOTE 2017-2018 Survey was a nationally representative survey of the Finnish population aged 20 years and over. The sampling frame was the Population Register of Statistics Finland [19]. The survey was based on random sampling. In 2017, 3300 people were invited to participate from 13 research areas (2300 adults aged 20-74 and 1000 adults aged 75+). In 2018, 500-3000 people were invited to participate from the same areas (adults over 75+ were oversampled). Data was collected between October 2, 2017, and March 3, 2018. Participants received a self-administered questionnaire in Finnish, Swedish, English, and Russian, which could be returned in paper or filled in electronically. The participation rate was 43%, resulting in an analytical sample of 26422 participants [19].

FinSote 2019 was a nationally representative survey of the Finnish population aged 20 and over, which was implemented in conjunction with the European Health Information Survey (EHIS) round 3. The sampling frame was the Population Register of Statistics Finland. The survey was based on a random sample of 15000 individuals. Participants received a self-administered questionnaire available in Finnish, Swedish and English, which could be returned in paper or filled in electronically. The participation rate was 44%, resulting in an analytical sample of ~6251 participants [20].

FinSote 2020 is a nationally representative survey of the Finnish population aged 20 and over. The sampling frame was the Digital and Population Data Services Agency, created in Jan 2020 after the merge between the Population Register of Statistics Finland and local register offices [21]. The survey is based on a random sample of 67000 individuals. Data collection started on September 14, 2020, and finished on February 8, 2021. Participants received a self-administered questionnaire in Finnish, Swedish, English, and Russian, which could be returned in paper or filled in electronically. The analytical sample is 28199 participants, with a participation rate of approximately 42,1%.

Exposure variables

The exposure of interest is tobacco use. FinSote 2018 and 2020 surveys have different modules for people aged 20-54, 55-74 and 75+. FinSote 2019 has different modules for those aged 15-54 and 55+. For all surveys, we will create a categorical variable for smoking with the following categories: never smokers, former smokers, occasional smokers, and daily smokers. For FinSote 2018 and 2020, we will additionally assess the use of snus, electronic cigarettes with and without nicotine and nicotine replacement therapy products. For all these additional indicators we will construct a categorical variable with the following categories: never users, former users, occasional users and daily users.

Confounders

Given that adjusting for controls that lie in the causal pathway between smoking and SARS-CoV-2 infection can lead to collider bias (see for example [2, 14, 22]), we will adjust for covariates that causally precede the exposure and are associated with the outcome [23]. Based on the directed acyclic graph shown in Figure 1 below, we will adjust for sex, age, marital status, years of education, mother tongue, and participation in social activities.

We will define marital status as those married or in a registered relationship or cohabiting versus those separated or divorced, widowed or single. We will measure years of education as the number of years a person has attended school or studied full time altogether. We will obtain information on the participant's mother tongue from national registries. We will categorize the mother tongue into Finnish, Swedish and others. We will measure participation with a question about participation in the activities of any club, association, hobby group or religious or spiritual community. We will categorise participation into the following: no participation, occasional and active.

In addition, participants living in different regions in Finland have varying risks of SARS-CoV-2 infection due to geographical variation in viral spread and diverging public health and social measures. Different prevalence of smoking by regions in Finland might confound the effect of smoking on the risk of SARS-CoV-2 infections. We will include regional fixed effects to account for these variations.

Collider bias

We will examine whether inducing potential collider bias (M-bias) by adjusting for other behavioural risk factors could explain earlier results [14, 15, 24, 25]. Other potential colliders, such as chronic conditions caused by tobacco use (e.g., chronic bronchitis) and SARS-CoV-2 testing, or hospitalizations are not possible due to lack of data.

Therefore, we will test the effect of collider bias due to alcohol use, physical activity and obesity. We will measure the volume of alcohol used by questions on the quantity of beer, wine and other alcoholic beverages consumed on a typical occasion and frequency of consumption. We will calculate the weekly grams of pure alcohol consumed by converting the alcohol servings consumed into grams of pure alcohol and multiplying them by the frequency of alcohol use (in a year). The result will be divided by 52 to obtain a measure of weekly alcohol use. We will measure leisure-time physical activity with a question on how often the participant engages in leisure exercise for a period of at least 30 minutes. We will categorise leisure-time physical activity into the following categories: daily; 4-6 times a week; 3 times a week; 2 times a week; once a week; 2-3 times a month; and a few times a year or less/I cannot exercise because of an illness or injury. We will calculate body mass index as the self-reported weight (in kg) divided by height (in m) squared. We will create a categorical variable using the classification of the World Health Organization: <18.5 underweight, 18.5–24.9 normal, 25–29.9 overweight, ≥30 obesity [26].

Outcomes

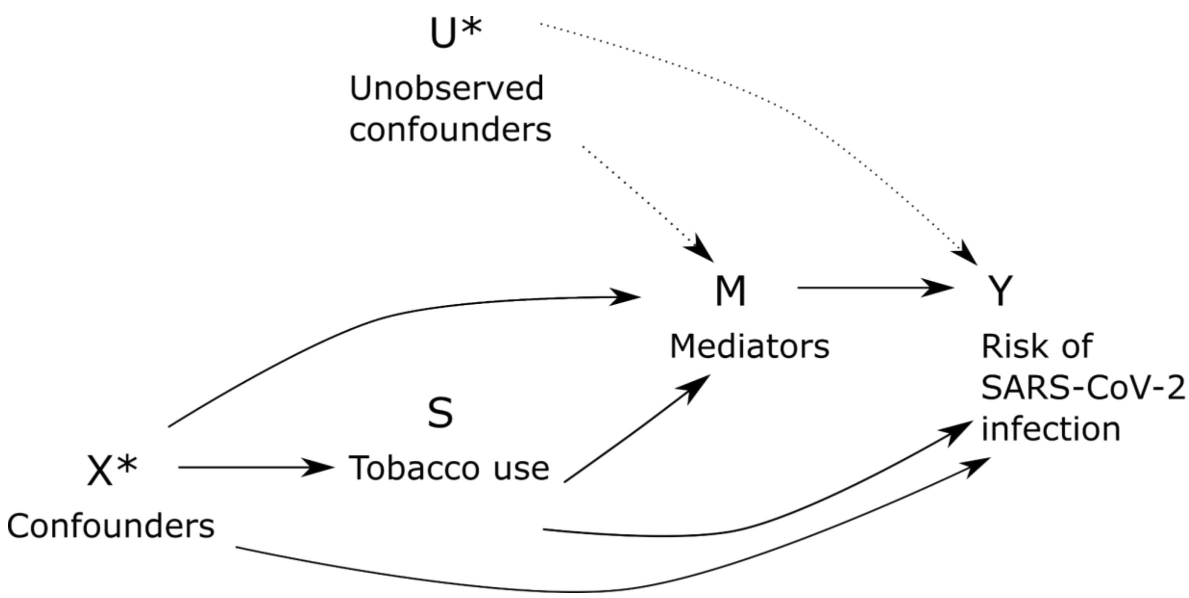
The primary outcome will be SARS-CoV-2 infection. We will obtain SARS-CoV-2 infection data from the Finnish National Infectious Disease Register maintained by the Finnish Institute for Health and Welfare. We will consider SARS-CoV-2 infections those cases with a positive RT-PCR and COVID-19 syndrome diagnosed by a physician [27]. Testing of SARS-CoV-2 is free in Finland. Coverage has been extensive and the total number of tests exceeds 5.0 million by June 2, 2021 [28].

Identification strategy

The study is observational, and we do not have a source of exogenous variation to obtain causal estimates. We will use the conditional independence assumption to approximate causal estimates [29, 30], as well as correctly defining confounders to prevent collider bias.

Figure 1 shows the directed acyclic graph for the study. X^* denotes a vector of confounders, S denotes the exposure to tobacco and Y the outcome of SARS-CoV-2 infection. The identifying assumption of conditional independence states that after conditioning on a set of observable covariates (i.e. confounders X^*), exposure to tobacco is independent of potential outcomes. In other words, after controlling for confounders, exposure to tobacco is assumed to be randomly assigned. This is a strong assumption, as there might be unobserved factors U leading to residual confounding. In this specific case, for example, certain personality traits (e.g. lower risk aversion) could increase the risk of tobacco use as well as be associated with higher risk of SARS-CoV-2 due to lower adherence to social distancing restrictions. Other unobserved factors could be religion or certain hobbies (e.g. singing in a choir), that might be associated with lower tobacco use but higher risk of SARS-CoV-2 infection.

Figure 1. Directed acyclic graph of the study



Note: Confounders X* in this study are sex, age, marital status and years of education. Unobserved confounders U* include personality traits, religion, or ethnic background. Mediators and potential colliders M include behavioural risk factors (observed), and other unobserved factors, such as SARS-CoV-2 testing or chronic health conditions caused by tobacco use.

Potential alternative identification strategies using, for example, instrumental variable design will Mendelian randomization be considered in further studies [31-33]. As a postal survey, FinSote does not have biological samples, but other datasets administered by the Finnish Institute of Health and Welfare have data on single nucleotide polymorphisms (SNPs) and could potentially be used to provide more robust causal estimates.

Statistical analyses

Given that the data is right censored, i.e. we are not able to observe the outcome in all participants in a given time frame due to lost-to-follow-up or because the event has not occurred during the study period, we will model it as time-to-event data.

We will use Cox proportional hazards models to estimate the risk of SARS-CoV-2 infections. Regression estimates will be presented as hazard ratios with 95% confidence intervals. We will use days in study as the timescale [34]. We will assess the proportional hazard assumption for each exposure covariate by visual inspection of plotted scaled Schoenfeld residuals against time and also testing the null hypothesis of zero slope for covariates and globally [35, 36].

We will run the following Cox model [37]:

$$(1) \qquad log\lambda_i = log\lambda_0(t) + \beta_1 S_i + \beta_x X_i^* + \rho_1 R_a$$

where λ_i is the hazard of being infected by SARS-CoV-2 for individual i; $\lambda_0(t)$ is the baseline hazard; $\beta_1$is the coefficient of interest for exposure to tobacco S; a vector of covariates X^* (i.e. sex, age, marital status, years of education); and fixed effects $\rho_1 R_a$ for region a.

We will test non-linearity in the association between age and the outcome and years of education and the outcome comparing the linear model with penalized smoothing splines using a likelihood ratio test [38].

We will test the possibility of collider bias due to behavioural risk factors by assessing the change in the coefficient β_1 after adjusting for volume of alcohol used, leisure-time physical activity and obesity.

We will use R version 3.6.3 for all analyses. We will use the *survival* package to run the Cox proportional hazards model.

References

1. Dong E., Du H., Gardner L. An interactive web-based dashboard to track COVID-19 in real time, Lancet Infectious Diseases 2020: 20: 533 - 534.
2. Williamson E. J., Walker A. J., Bhaskaran K., Bacon S., Bates C., Morton C. E. et al. Factors associated with COVID-19-related death using OpenSAFELY, Nature 2020: 584: 430-436.
3. Vardavas C. I., Nikitara K. COVID-19 and smoking: A systematic review of the evidence, Tob Induc Dis 2020: 18: 20.
4. Van Westen-Lagerweij N. A., Meijer E., Meeuwsen E. G., Chavannes N. H., Willemsen M. C., Croes E. A. Are smokers protected against SARS-CoV-2 infection (COVID-19)? The origins of the myth, NPJ Prim Care Respir Med 2021: 31: 10.
5. Alla F., Berlin I., Nguyen-Thanh V., Guignard R., Pasquereau A., Quelet S. et al. Tobacco and COVID-19: a crisis within a crisis?, Can J Public Health 2020: 111: 995-999.
6. Berlin I., Thomas D., Le Faou A.-L., Cornuz J. COVID-19 and Smoking, Nicotine & Tobacco Research 2020: 22: 1650-1652.
7. Edwards R., Munafò M. COVID-19 and Tobacco: More Questions Than Answers, Nicotine Tob Res 2020: 22: 1644-1645.
8. Moore S., Teixeira A., Stewart S. Effect of Network Social Capital on the Chances of Smoking Relapse: A Two-Year Follow-up Study of Urban-Dwelling Adults, American Journal of Public Health 2014: 104: e72-e76.
9. Setti L., Passarini F., De Gennaro G., Barbieri P., Perrone M. G., Borelli M. et al. Airborne Transmission Route of COVID-19: Why 2 Meters/6 Feet of Inter-Personal Distance Could Not Be Enough, International Journal of Environmental Research and Public Health 2020: 17: 2932.
10. Engin A. B., Engin E. D., Engin A. Two important controversial risk factors in SARS-CoV-2 infection: Obesity and smoking, Environmental toxicology and pharmacology 2020: 78: 103411-103411.
11. Cai G., Bossé Y., Xiao F., Kheradmand F., Amos C. I. Tobacco Smoking Increases the Lung Gene Expression of ACE2, the Receptor of SARS-CoV-2, Am J Respir Crit Care Med 2020: 201: 1557-1559.
12. Haddad C., Bou Malhab S., Sacre H., Salameh P. Smoking and COVID-19: A Scoping Review, Tob Use Insights 2021: 14: 1179173x21994612.
13. Jiménez-Ruiz C. A., López-Padilla D., Alonso-Arroyo A., Aleixandre-Benavent R., Solano-Reina S., De Granda-Orive J. I. COVID-19 and Smoking: A Systematic Review and Meta-Analysis of the Evidence, Arch Bronconeumol 2021: 57 Suppl 1: 21-34.
14. Griffith G. J., Morris T. T., Tudball M. J., Herbert A., Mancano G., Pike L. et al. Collider bias undermines our understanding of COVID-19 disease risk and severity, Nature Communications 2020: 11: 5749.
15. Tattan-Birch H., Marsden J., West R., Gage S. H. Assessing and addressing collider bias in addiction research: the curious case of smoking and COVID-19, Addiction 2021: 116: 982-984.
16. Simons D., Shahab L., Brown J., Perski O. The association of smoking status with SARS-CoV-2 infection, hospitalization and mortality from COVID-19: a living rapid evidence review with Bayesian meta-analyses (version 7), Addiction 2020.
17. Bar-Zeev Y. Commentary on Simons et al. Public health implications of the suggested association between nicotine, smoking and infection with SARS-CoV-2, Addiction 2021: 116: 1369-1370.
18. Vandenbroucke J. P., Von Elm E., Altman D. G., Gøtzsche P. C., Mulrow C. D., Pocock S. J. et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): Explanation and Elaboration, PLoS Med 2007: 4: e297.

19. Pentala-Nikulainen O., Koskela T., Parikka S., Aalto A.-M., Muuri A. Alueelliset erot aikuisväestön palvelukokemuksissa ja hyvinvoinnissa – FinSote 2018. Helsinki, Finland: Finnish Institute for Health and Welfare; 2018.
20. Parikka S., Koskela T., Ikonen J., Kilpeläinen H., Tolonen H. Implementation of the European health interview survey into the National FinSote survey in Finland, European Journal of Public Health 2020: 30.
21. Digital and Population Data Services Agency. Digital and Population Data Services Agency: toward a smooth and functioning Finland. 2021. Available at: https://dvv.fi/en/individuals (accessed March 30, 2021)
22. Greenland S., Pearl J., Robins J. M. Causal diagrams for epidemiologic research, Epidemiology 1999: 10: 37-48.
23. Hernán M. A., Hernández-Díaz S., Werler M. M., Mitchell A. A. Causal knowledge as a prerequisite for confounding evaluation: an application to birth defects epidemiology, Am J Epidemiol 2002: 155: 176-184.
24. Pearce N., Lawlor D. A. Causal inference-so much more than statistics, International journal of epidemiology 2016: 45: 1895-1903.
25. Sacerdote C., Ricceri F., Rolandsson O., Baldi I., Chirlaque M.-D., Feskens E. et al. Lower educational level is a predictor of incident type 2 diabetes in European countries: The EPIC-InterAct study, International Journal of Epidemiology 2012: 41: 1162-1173.
26. World Health Organization. Physical status: the use and interpretation of anthropometry. Report of a WHO Expert Committee, World Health Organ Tech Rep Ser 1995: 854: 1-452.
27. Finnish Institute for Health and Welfare. Compilation of statistics for COVID-19 cases. 2021. Available at: https://thl.fi/en/web/infectious-diseases-and-vaccinations/what-s-new/coronavirus-covid-19-latest-updates/situation-update-on-coronavirus/compilation-of-statistics-for-covid-19-cases (accessed March 30, 2021)
28. Finnish Institute for Health and Welfare. Situation update on coronavirus. 2021. Available at: https://thl.fi/en/web/infectious-diseases-and-vaccinations/what-s-new/coronavirus-covid-19-latest-updates/situation-update-on-coronavirus (accessed March 30, 2021)
29. Lechner M. Identification and estimation of causal effects of multiple treatments under the conditional independence assumption, Heidelberg: Physica-Verlag HD; 2001, p. 43-58.
30. Dale S. B., Krueger A. B. Estimating the Payoff to Attending a More Selective College: An Application of Selection on Observables and Unobservables*, The Quarterly Journal of Economics 2002: 117: 1491-1527.
31. Larsson S. C., Burgess S., Michaëlsson K. Smoking and stroke: A mendelian randomization study, Annals of Neurology 2019: 86: 468-471.
32. Taylor A. E., Fluharty M. E., Bjørngaard J. H., Gabrielsen M. E., Skorpen F., Marioni R. E. et al. Investigating the possible causal association of smoking with depression and anxiety using Mendelian randomisation meta-analysis: the CARTA consortium, BMJ Open 2014: 4: e006141.
33. Wehby G. L., Fletcher J. M., Lehrer S. F., Moreno L. M., Murray J. C., Wilcox A. et al. A genetic instrumental variables analysis of the effects of prenatal smoking on birth weight: evidence from two samples, Biodemography and social biology 2011: 57: 3-32.
34. Korn E. L., Graubard B. I., Midthune D. Time-to-event analysis of longitudinal follow-up of a survey: choice of the time-scale, Am J Epidemiol 1997: 145: 72-80.
35. Zhang Z., Reinikainen J., Adeleke K. A., Pieterse M. E., Groothuis-Oudshoorn C. G. M. Time-varying covariates and coefficients in Cox regression models, Ann Transl Med 2018: 6: 121.
36. Harrell F. Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis Switzerland: Springer; 2015.

37. Huang Y.-T., Yang H.-I. Causal Mediation Analysis of Survival Outcome with Multiple Mediators, Epidemiology (Cambridge, Mass) 2017: 28: 370-378.
38. Ramsay J. O., Heckman N., Silverman B. W. Spline smoothing with model-based penalties, Behav Res Methods Instrum Comput 1997: 29: 99-106.