

# TOBACCO ABUSE IN EUROPE

Tutorial 3 Group 3,  
Gyan Gadjoe (2807574),  
Jason Purnama (2827689),  
Josie Chen (2801717),  
Mijke Gerritsen (2824547),  
Milica Stankovic(2847467)  
Veronika Vozarova(2847711)  
Noel Tripathi (2871659)

Chantal Schouwenaar,  
Jack Fitzgerald,

Programming for Economists  
Vrije Universiteit Amsterdam

June 2025

# Part 1 - Identify a Social Problem

## 1.1 Describe the Social Problem

Include the following:

- Why is this relevant?

People who smoke still pose a major threat to public health and society in Europe. They kill many people who could have been saved thanks to tobacco use. From 2019 to 2022, smoking was always connected to higher death rates in all EU member states. More than 850,000 people died in the EU in 2019 because they were actively smoking, and another 73,000 died because they were exposed to secondhand smoke (European Commission, 2021). During this time, countries like Hungary, Bulgaria, and Latvia had death rates of more than 250 deaths per 100,000 inhabitants from illnesses caused by smoking (OECD, 2023).

Additionally, smoking is bad for your health and costs healthcare systems money, makes people less productive, and makes health disparities worse. According to the OECD (2023), the number of people aged 15 and up who smoked every day in Europe stayed high from 2019 to 2022. It was less than 10% in some Nordic countries and more than 25% in Eastern Europe.

The effects last a long time and are deadly: smokers lose an average of 10 to 14 years of life and are more likely to have lung cancer, heart disease, and respiratory diseases (WHO, 2023). Millions of people who don't smoke are still getting sick from secondhand smoke, especially children, the old, and low-income families (CDC, 2025). This means that smoking is not just bad for your health, but it is also a long-term societal problem that needs help from other countries' policies, health education, and community assistance programs.

## Part 2 - Data Sourcing

### 2.1 Load in the data

The analysis was made up of two datasets, specifically two data sets from the OECD Health statistics

1. **Tobacco consumption data** [OECDnew.csv](#): This reports the average cigarettes consumed per smoker per day for the OECD countries from the year 2019 up until the year 2022.
2. **Avoidable Mortality data** [OECD.avoidable death.csv](#): This file includes country-level, avoidable, mortality rate rates (per 100,000 inhabitants) and this is from the year 2019 up until the year 2022.

For both of these data sets were loaded using `read_csv()` and we cleaned to retain the needed variables, for example: country codes, years, mortality, rate, and tobacco consumption. And other missing values and non-applicable data or filtered out.

### 2.2 Provide a short summary of the dataset(s)

#### Data set 1: Avoidable mortality

**Rows:** this data has about 500 rows of data including 30 countries across 4 years from the year 2019 up until 2022, which would be 120 expected rows per variable.

#### Key Variables :

**mortality\_rate:** Age standardized deaths per 100,000

**Year:** 2019-2022

#### Data set 2 :tobacco consumption.

**Rows:** this data set includes about 300 rows of data, including 22 countries for 4 years from the year 2019 up until the year 2022, which is about 88 rows of data per variable.

#### Key variables:

**Cigarettes\_per\_day:** the average number of cigarettes consumed daily by each smoker in the population.

**Years:** 2019-2023

**obs\_status:** Only “A” (actual) values kept

## 2.3 Describe the type of variables included

### **Data set 1: avoidable mortality**

For this data set there are 4 different variables that where applicable: Country\_code, Country, year and mortality\_rate.

**Country\_code:** The unit is ISO3, which is a standardized country identifier, and the timeframe for this is 2019 to 2022 and the limitation that are present was non-OECD countries were excluded.

**Country:** The unit used for this is just country name, which is just the official name of the country for example “ Netherlands”, and the timeframe for this is 2019 to 2022 and there were no real limitations.

**year:** For this type of variable, it was an integer and the units used were just calendar years, to be exact January 1 to December 31 and the timeframe for this is 2019 to 2022. and there were certain limitations here for example. A few countries did not have data in the year 2022.

**mortality\_rate:** For this type of variable, the unit used was deaths per 100,000 inhabitants and this was numeric. The timeframe for this was 2019 to 2022, and there were no real limitations.

### **Data set 2: tobacco consumption**

For this data set, 4 variables were used, which include country\_code, Year, cigar\_per\_day and obs\_status.

**Country\_code:** The unit for this ISO3, which is a standardized country identifier, and the time duration for This is 2019 to 2022. A limitation that emerged is there are only 22/30 countries available.

**Year:** For this variable, the units used were calendar years, so January 1 - December 31 and the time duration For this is 2019-2022. There were no limitations for this variable.

**Cigarettes\_per\_day:** This is for the average daily consumption per smoker, and the time duration for this was From 2019 to 2022 and the one limitation is that the daily consumption per smoker is survey based, so this would exclude unreported tobacco.

**Obs\_status:** This is a metadata flag for data set tobacco consumption, and mostly gives the quality and measurement type of each data point. Its filtered by “ obs\_status == A” so the values are directly measured as opposed To “E” which is estimated. Time duration for this is 2019-2022.

## Part 3 - Quantifying

### 3.1 Data cleaning

For both datasets (OECDnew.csv and OECD.avoidable death.csv), the following cleaning steps were performed:

- Necessary column was selected: Country, Year, Value
- Filtered for the relevant indicators: Tobacco consumption per smoker and avoidable mortality rate.
- Removed any NA values to ensure clean, complete datasets.
- Merged the two datasets using the Country and Year variables for analysis.

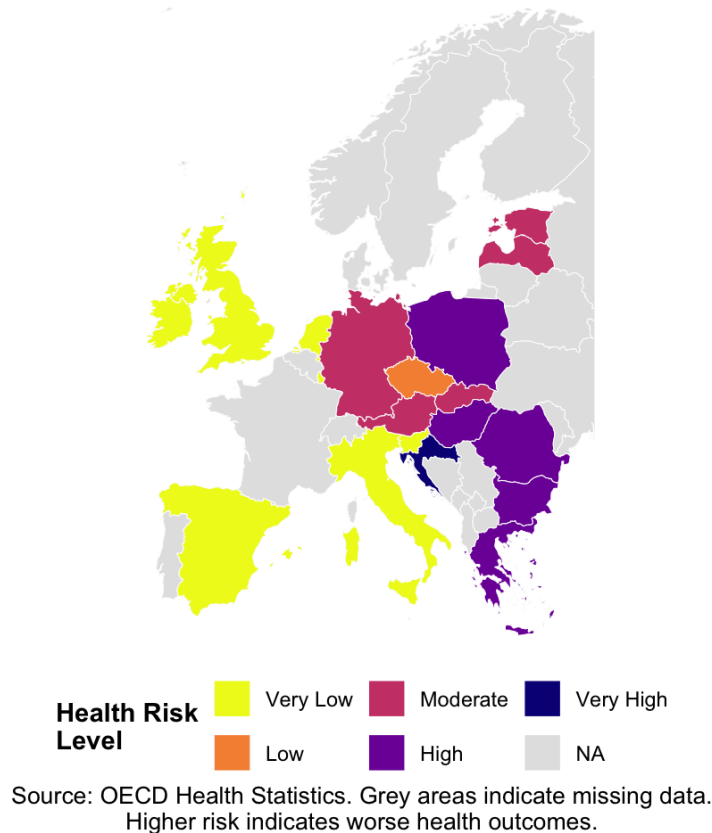
### 3.2 Generate necessary variables

To measure the combined risk level of tobacco use and preventable mortality, a **Health Risk Index** (HRI) was developed by:

- Using z-scores to standardise the variables related to tobacco use and preventable mortality. The HRI is calculated by averaging these standardised scores.
- Based on the HRI's quartiles, countries were divided into four risk categories: Very High, High, Moderate, and Low Risk.

## Health Risk Index Across European Countries

Combined indicator of avoidable mortality and tobacco consumption (2019-2022)



### 3.3 Visualize temporal variation

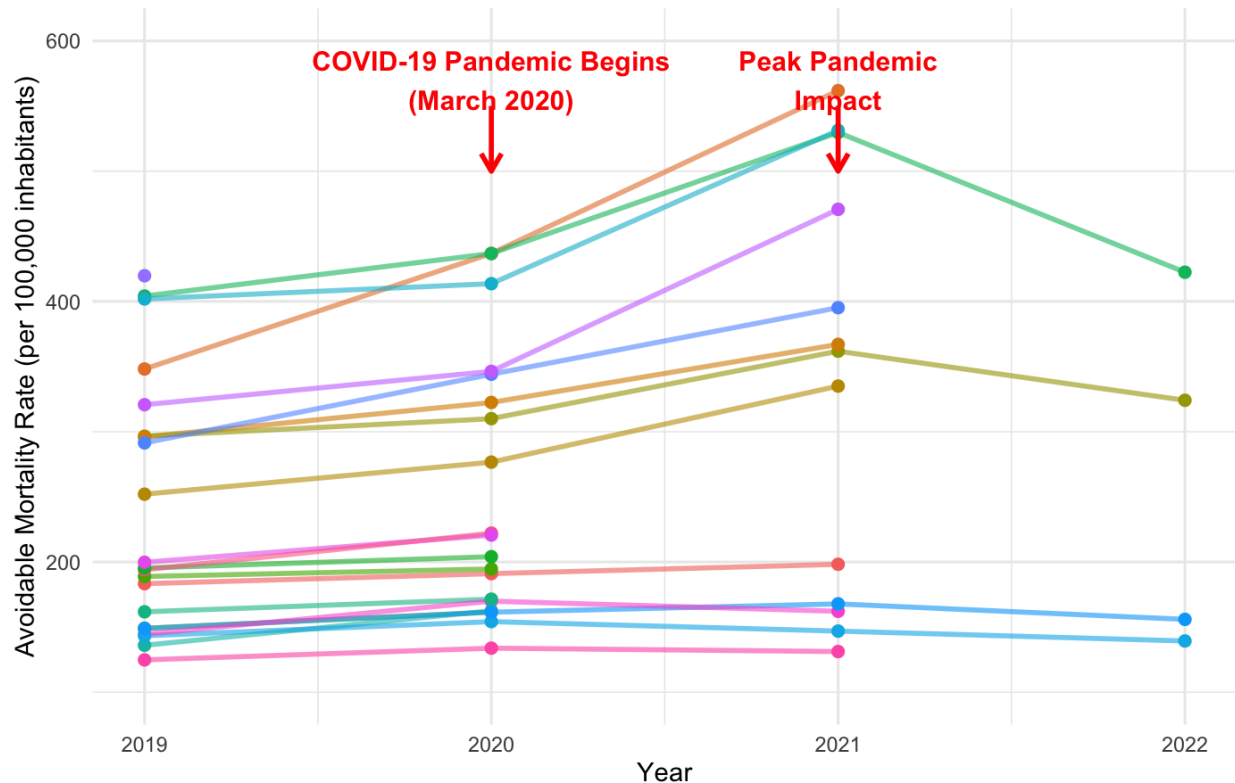
The average preventable mortality rates and tobacco use in every nation from 2019 to 2022 were displayed in a time series line plot.

#### Why this matters:

This graphic shows how smoking behaviour and its possible health effects have changed over time. The plot helps in identifying any temporal pattern, such as the mortality peak in 2020–2021, which may be connected to the public health effects of the COVID-19 pandemic.

### COVID-19 Impact on Avoidable Mortality (2019-2022)

Sharp increases observed during pandemic years across European countries



Source: OECD Health Statistics. Red shaded area highlights pandemic period.

### 3.4 Visualize spatial variation

This identifies geographical differences in health risks related to smoking. For example, Eastern and Southern European countries tend to be in higher risk categories, highlighting regions needing stronger tobacco control and health interventions.

### 3.5 Visualize sub-population variation

not applicable for us.

Avoidable mortality rates in high-risk nations were separated into three groups: above median, below median, and middle. This showed that, compared to other countries, those above the median consistently had worse health outcomes and higher tobacco use.

### 3.6 Event analysis

A scatter plot was created to examine the relationship between:

- Tobacco consumption per smoker per day and avoidable mortality rate.

To analyze the relationship between preventable death rates (per 100,000) and tobacco consumption (cigarettes per smoker per day) in European nations, highlighting the ways in which this relationship changes depending on the level of national health risk.

The plot shows that: (1) there is an upward trend between tobacco use and mortality, which is largest in high-risk countries ( $r = 0.60$ ); (2) the connections are weaker in low-risk countries, indicating reasons other than tobacco use; and (3) smoking-related policy changes could reduce mortality differences, particularly in high-burden regions. This supports WHO equity goals by highlighting tobacco's contribution to avoidable deaths.



## Part 4 - Discussion

### 4.1 Discuss your findings

From 2019 to 2022, the OECD will study health statistics to see how smoking causes deaths that aren't necessary in different countries. This is done with the help of a joint Health Risk Index (HRI) that displays all the health risks that the community faces. As a result of standardizing and average death rates and the number of cigarettes people smoke every day, the HRI was able to divide countries into four groups: Very High Risk, High Risk, Moderate Risk, and Low Risk. It was found that countries with higher HRIs have worse health effects. It seems that both the number of smokers and the number of deaths are important in figuring out a country's health risk profile. During the COVID-19 outbreak, there was a clear and large rise in deaths that were not necessary. This was the worst between 2020 and 2021. It looked like things were getting better by 2022. People all over the world have been touched by the epidemic, and these trends show how different countries have dealt with it.

Spatial analysis showed that different areas have different health. Usually, Eastern and Southern Europe had more danger, while Western and Northern Europe did better. Another study that looked at groups of people found that smoking was more strongly linked to needless deaths in high-risk countries than in low-risk countries. This means that smoking seems to have a bigger effect when health services are already under a lot of stress. The event analysis showed that 2020 and 2021 were years with a lot of deaths, which was in line with the world COVID-19 disaster. Overall, this joint study helps us figure out which groups of people are most at risk, pick the first public health steps, and make rules that will save lives by changing how people act and how things are set up.

## **Part 5 - Reproducibility**

### 5.1 Github repository link

<https://github.com/Subjectbutton53/substance-abuse-3-3>

## 5.2 Reference list

Centers for Disease Control and Prevention. (2025, January 31). *Health problems caused by secondhand smoke*. <https://www.cdc.gov/tobacco/secondhand-smoke/health.html>

European Commission. (2021). *Mortality attributable to tobacco smoking in EU Member States (2019 data)*. Health Promotion Knowledge Gateway. [https://knowledge4policy.ec.europa.eu/health-promotion-knowledge-gateway/smoking-deaths\\_en?](https://knowledge4policy.ec.europa.eu/health-promotion-knowledge-gateway/smoking-deaths_en?)

Organisation for Economic Co-operation and Development. (2023). *Daily smoking among adults aged 15 and over (2011–2021)*. <https://doi.org/10.1787/7a7afb35-en>

OECD. (n.d.). *Health; living conditions – Health, living time-common health* [Dataset]. OECD Data Explorer. Retrieved June 24, 2025, from [https://data-explorer.oecd.org/vis?lc=en&df\[ds\]=dsDisseminateFinalDMZ&df\[id\]=DSD\\_HEALTH\\_LVNG%40DF\\_HEALTH\\_LVNG\\_TC&df\[ag\]=OECD.ELS.HD&dq=TUR%2BCHE%2BSWE%2BESP%2BSVN%2BSVK%2BPRT%2BPOL%2BNOR%2BNLD%2BLUX%2BLTU%2BLVA%2BITA%2BISL%2BHUN%2BGR%2BDEU%2BFRA%2BFIN%2BEST%2BDNK%2BCZE%2BCAN%2BAUT.A.TBC.T.T.CG\\_SM\\_D.&pd=2019%2C2022&to\[TIME\\_PERIOD\]=false&vw=tb](https://data-explorer.oecd.org/vis?lc=en&df[ds]=dsDisseminateFinalDMZ&df[id]=DSD_HEALTH_LVNG%40DF_HEALTH_LVNG_TC&df[ag]=OECD.ELS.HD&dq=TUR%2BCHE%2BSWE%2BESP%2BSVN%2BSVK%2BPRT%2BPOL%2BNOR%2BNLD%2BLUX%2BLTU%2BLVA%2BITA%2BISL%2BHUN%2BGR%2BDEU%2BFRA%2BFIN%2BEST%2BDNK%2BCZE%2BCAN%2BAUT.A.TBC.T.T.CG_SM_D.&pd=2019%2C2022&to[TIME_PERIOD]=false&vw=tb)

OECD. (n.d.). *Health status indicators by country, 2019–2022* [Dataset]. OECD Data Explorer. Retrieved June 24, 2025, from [https://data-explorer.oecd.org/vis?lc=en&fs\[o\]=Topic%2C1%7CHealth%23HEA%23%7CHealth%20status%23HEA\\_STA%23&pg=0&fc=Topic&bp=true&snb=17&vw=tb&df\[ds\]=dsDisseminateFinalDMZ&df\[id\]=DSD\\_HEALTH\\_STAT%40DF\\_COM&df\[ag\]=OECD.ELS.HD&df\[vs\]=1.0&dq=GBR%2BROU%2BHRV%2BBGR%2BTUR%2BCHE%2BSWE%2BESP%2BSVN%2BSVK%2BPRT%2BPOL%2BNLD%2BLUX%2BLTU%2BLVA%2BITA%2BISL%2BHUN%2BGR%2BDEU%2BFRA%2BFIN%2BEST%2BDNK%2BCZE%2BAUT%2BBEL.A....T..T.STANDARD....&pd=2019%2C2022&to\[TIME\\_PERIOD\]=false](https://data-explorer.oecd.org/vis?lc=en&fs[o]=Topic%2C1%7CHealth%23HEA%23%7CHealth%20status%23HEA_STA%23&pg=0&fc=Topic&bp=true&snb=17&vw=tb&df[ds]=dsDisseminateFinalDMZ&df[id]=DSD_HEALTH_STAT%40DF_COM&df[ag]=OECD.ELS.HD&df[vs]=1.0&dq=GBR%2BROU%2BHRV%2BBGR%2BTUR%2BCHE%2BSWE%2BESP%2BSVN%2BSVK%2BPRT%2BPOL%2BNLD%2BLUX%2BLTU%2BLVA%2BITA%2BISL%2BHUN%2BGR%2BDEU%2BFRA%2BFIN%2BEST%2BDNK%2BCZE%2BAUT%2BBEL.A....T..T.STANDARD....&pd=2019%2C2022&to[TIME_PERIOD]=false)

World Health Organization. (2023). *WHO report on the global tobacco epidemic, 2023: Protect people from tobacco smoke*. <https://www.who.int/publications/i/item/9789240077164>