Drugs in municipal wastewater in selected European cities
OPEN SOURCE DATA PROJECT

Viktoria Dyk

18.06.2022

—

Data Analytics

EXECUTIVE SUMMARY

Wastewater analysis is a rapidly developing scientific discipline with the potential for monitoring illicit drug use. The European Monitoring Centre for Drugs and Drug Addiction (EMCDDA) works closely with the Sewage analysis group (SCORE) which has been collecting data indicating the amount of drug residues in watstewater in major European cities since 2011. We will explore geographical and temporal trends in illicit drug use covering cannabis, cocaine, amphetamine, methamphetamine and MDMA (ecstasy).

DATA SOURCE

Two dataset has been published by the EMCDDA, one about the measured values indicating the amount of drug residues between 2011-2021 and one about the sites (wastewater treatment centers) where the measurements have been conducted. Link for the datesets are here

The datasets are considered to be trustworthy as it is administrative data collected by the leading authority on illicit drugs in the European Union – provided to them directly by the wastewater treatment centers.

Merging the two datasets results in a dataset which fulfills the project requirements and facilitates spatial and time series analysis.

LIMITATIONS AND ETHICS

- The measurements cover 75 cities and 23 countries in the European Union, Norway and Turkey between 2011-2021 but for some years and sites measurements are missing.
- Some values are below the method limit of quantification: these are indicated as zero in the dataset.
- Population data for the cities is available only for 2021.

Wastewater-based epidemiology is not conducted on individuals and therefore considered generally as low-risk in terms of ethics. In our case the measurements were also not conducted on small communities and no individual groups can be identified.

DATA COLLECTION

A standard protocol and a common quality control exercise were used in all locations to compare illicit drug loads in Europe over a one-week period during 10 consecutive years. During the years more and more cities joined performing the standardized measurements.

In 2021, raw 24-hour composite samples were collected during a single week between March and May 2021. These samples were analysed for the urinary biomarkers of the parent drug.

DATA CLEANING STEPS

After merging the two datasets on the SiteID, the following data cleaning steps have been conducted:

- dropping the columns that are not needed for the analysis
- renaming some columns to have a consistent naming convention
- changing datatype for column “latitude” (from string to floating)

DATA CONSISTENCY CHECKS

- preparing descriptive statistics (see script) – no inconsistencies detected
- preparing frequency tables for qualitative variables: for cities where multiple treatment plants are available, the number of plants are indicated after the city name*
- looking for mixed datatypes: none found
- looking for missing values: some measurement values missing – we are aware of this data limitation and will leave them missing
- looking for duplicates: none found

*Note: not sure if this will cause an issue in the spatial analysis therefore didnt change the naming in these cases now

DATA PROFILE

Number of columns and rows of the final dataset: 2569 x 18

Column details in table format:

Column	Column description	Datatype	Structured	Time Variant
year	Year of measurement	Quantitative, Discrete	yes	no
metabolite	type of drug residues	Qualitative, Nominal	yes	no
SiteID	ID of treatment centre where the measurement was performed	Qualitative, Ordinal	yes	no
country	country code of measurement	Qualitative, Nominal	yes	no
city	city of measurement	Qualitative, Nominal	yes	no
Wednesday	measured residue in mg/1000 person on this day	Quantitative, Continous	yes	yes
Thursday	measured residue in mg/1000 person on this day	Quantitative, Continous	yes	yes
Friday	measured residue in mg/1000 person on this day	Quantitative, Continous	yes	yes
Saturday	measured residue in mg/1000 person on this day	Quantitative, Continous	yes	yes
Sunday	measured residue in mg/1000 person on this day	Quantitative, Continous	yes	yes
Monday	measured residue in mg/1000 person on this day	Quantitative, Continous	yes	yes
Tuesday	measured residue in mg/1000 person on this day	Quantitative, Continous	yes	yes
Weekday mean	averages of measured drug residues on Tuesdays, Wednesdays and Thursdays	Quantitative, Continous	yes	yes
Weekend mean	averages of measured drug residues on Fridays, Saturdays, Sundays and Mondays	Quantitative, Continous	yes	yes
Daily mean	average of measured drug residues on all days of the week	Quantitative, Continous	yes	yes
latitude	geographic coordinates of the wastewater center	Quantitative, Continous	yes	no
longitude				
population	population of the city in 2021	Quantitative, Discrete	yes	no

QUESTIONS TO EXPLORE

- Which drugs are the most used in which area of Europe?
- What geographical and temporal trends can be observed?
- Is there a connection between drug usage and city size / drug usage and country's wealth (GDP per capita)? *
- How did the pandemic affect the drug usage in 2020-2021 when social events and interactions were strongly limited?

*Note: I am planning to add country GDP per capita to the dataset for the next task where we are exploring relationships between variables