tinbergen institute

Temporal, Spatial, Economic and Crime Factors in Illicit Drug Usage across European Cities

Jacques J.F. Commandeur [1]
Suncica Vujic[2]
Siem Jan Koopman[1,3]
Barbara Kasprzyk-Hordern[4]

[1] *Faculty of Economics and Business Administration, VU University Amsterdam, the Netherlands;*
[2] *University of Antwerp, Belgium;*
[3] *Tinbergen Institute, the Netherlands;*
[4] *University of Bath, United Kingdom.*

Temporal, Spatial, Economic and Crime Factors in Illicit Drug Usage across European Cities

Jacques J.F. Commandeur[a], Sunčica Vujić[b],
Siem Jan Koopman[a,c,d], Barbara Kasprzyk-Hordern[e]*

[a] *Vrije Universiteit Amsterdam, The Netherlands*
[b] *University of Antwerp, Belgium* [c] *Tinbergen Institute, The Netherlands*
[d] *CREATES, Aarhus University, Denmark* [e] *University of Bath, U.K.*

Abstract

We analyze the illicit drug usage by inhabitants and visitors of European cities. Our statistical analyses are by means of linear mixed models. The data on illicit drug usage of cocaine, ecstasy, amphetamines, methamphetamines, and cannabis are collected through wastewater samples from the inlet of 21 sewage treatment plants spread over 11 European countries. The data set represents nineteen cities, services a population of approximately 15 million inhabitants and covers a one-week period in 2011. The patterns of illicit drug usage are examined with respect to temporal (daily) and spatial variations, as well as in relation to economic wealth (gross domestic product) and criminological (drug offences recorded by police) factors. In a joint statistical analysis, we find that cocaine and ecstasy are typically recreational drugs that are consumed during the weekend. Inhabitants of Western European countries consume more cocaine than inhabitants of Eastern European countries. This finding cannot be explained by political divisions between West and East. We also find evidence that higher usage of ecstasy is associated with medium-sized cities, economic prosperity, and a lower number of drug offences. On the other hand, higher usage of methamphetamine is associated with medium-sized cities and low economic wealth.

Keywords: Sewage biomarker analysis, ANOVA, Linear mixed models, Wastewater-based epidemiology.

---

*Corresponding author: Sunčica Vujić, Department of Economics, University of Antwerp, Prinsstraat 13, 2000 Antwerpen, Belgium. Tel: +32 (0)32654902 Email: `suncica.vujic@uantwerpen.be`. October 15, 2014.

1 Introduction

According to the European Monitoring Centre for Drugs and Drug Addiction, as reported in their 2011 report, 23.2% of European adults (15-64 years old) use cannabis, 4.3% use cocaine, 3.2% use ecstasy, and 3.8% use amphetamines. Although the historic trends show that drug usage appears to be relatively stable in Europe, it still poses significant challenges to public policies aimed at drug use detection. In particular, the new developments in the synthetic drugs market raise concern. The synthetic drugs are difficult to detect and are not subject to current controls. They are also not monitored by public health organizations for drug policies with objectives such as prevention of HIV/AIDS epidemics, overdose deaths and other drug-related health problems. Public policies aimed at reducing the number of offences related to synthetic drug use are still in their infancy.

This paper utilizes a novel data set to investigate the patterns of illicit drug usage across European cities. The basis of our data set is obtained from Thomas et al. (2012) who discuss various (technical) details of how the proxies of drug usage are constructed from wastewater data. We extend their data set in order to analyze the patterns of illicit drug usage with respect to temporal (daily) and spatial (geographical) variations, as well as in relation to economic (gross domestic product, GDP) and criminological (per capita number of drug offences recorded by the police) factors. The first contribution of the paper is to provide a rigorous statistical analysis for this novel data set from Thomas et al. (2012) on illicit drug usage across multiple European locations in 2011.

Wastewater-based epidemiology (WBE) enables retrieving epidemiological information via the analysis of human metabolic excretion products in wastewater. The WBE approach is based on the fact that human urinary biomarkers (such as drug metabolites) resulting from exposure to illicit drugs are collected and pooled by the sewage system. It provides evidence of the amount and type of drug to which a population has been exposed to; see the discussions in Thomas et al. (2012) and Ort et al. (2014)[1].

The data on illicit drug usage of cocaine, ecstasy, amphetamines, methamphetamines, and cannabis were collected through the WBE analysis applied to wastewater samples from the inlet of 21 sewage treatment plants spread over 11 European countries and 19 cities, covering a population of approximately 15 million inhabitants. In comparison to data based on urine testing and/or self-reported substance use, the data collected from sewage samples is typically more representative of the illicit drug usage of the population in a particular city. Data collected from surveys of individuals who agreed to participate in the survey or particular other categories of the population (for example adults, prisoners, juveniles, etc.) is prone to the self-reported

---

[1]More general information on wastewater-based epidemiology can be obtained from `http://score-cost.eu` and `http://sewprof-itn.eu`.

drug use bias. The WBA analysis based on wastewater data is obviously free from this bias. Furthermore, it can be obtained within shorter time frames than the survey data on drug use and can help in identifying a range of illicit drugs usage, see Ort et al. (2014). A clear overview of drug use in Europe is important because behavioural trends in drug use can be changing quickly. For example, a wider set of substances are used by the consumers (polydrug use) including the combination of illicit drugs with alcohol, medicine and various non-controlled substances. Also, the new psychoactive substances are difficult to detect and not subject to current controls. We therefore conclude that improving data sets and having more insights of drug usage are still of eminent importance.

Thomas et al. (2012) discuss in more detail how raw sewage samples were collected via a WBE analysis from the inlet of 21 sewage treatment plants spread over 19 European cities. Samples were collected from each location over seven consecutive days, starting on Wednesday March 9, 2011, and ending on Tuesday, March 16, 2011. From these samples daily loads relative to population (mg/day/1000 inhabitants) were estimated of the illicit usage of cocaine, ecstasy, amphetamines, methamphetamines, and cannabis. For each of these drugs, we obtained potentially a data set of 133 (that is 19 cities times 7 days) observations. For each drug separately, these data were then averaged over all seven sampling days in order to compare illicit drug usage across the nineteen European cities (irrespective of time). Furthermore, for each of these drugs the day-to-day variation of illicit usage was determined by calculating the median over all nineteen cities of the daily drug loads expressed as percentage of the measured total weekly load. The data clearly show that illicit cocaine and ecstasy usage during the weekend is significantly higher than during the week; see Thomas et al. (2012) who attribute this finding to the recreational usage of these substances.

The second contribution of our paper is to provide a more insightful analyis of the illicit drug use patterns across European cities by using advanced statistical methods which are based on linear mixed models; see the textbook treatments of Twisk (2013) and Stroup (2013). In this advanced statistical framework, for each drug, the nineteen cities with seven repeated observations are analyzed simultaneously.

Linear mixed models provide a framework in which the analysis of drug usage can be embedded into a larger societal context by introducing between-cities factors and covariates into the model. We further investigate the relationships between drug usage and economic and criminological aspects of the nineteen cities involved in the study. The gross domestic product (GDP) is used as a proxy for the economic prosperity in each country, while the per capita number of drug offences recorded by the police is used as a proxy either for drug related criminal activity in each country or for the success the police has in apprehending citizens involved in drug related criminal offences. Our third contribution is to provide a simultaneous analysis to the patterns of illicit drug use with respect to temporal (daily) and spatial (geographic)

variations, as well as in relation to economic (GDP) and criminological (per capita number of drug offences recorded by the police) factors. In effect our aim is to contribute to the current policy debates with respect to the reduction of the number of offences related to drug use.

The remaining part of the paper is organised as follows. In Section 2 we present the data. The empirical results of the analyses with these models are presented in Section 3. In Section 4 we discuss and review the results. The adopted methodology of linear mixed models is reviewed and discussed in the Appendix.

2   Data

In our study to illicit drug use by inhabitant and visitors of European cities, we consider 19 European cities where sewage biomarker data has been collected for the study of Thomas et al. (2012). The cities are Antwerp and Brussels (Belgium), Zagreb (Croatia), Budweis (Czech Republic), Helsinki and Turku (Finland), Paris (France), Milan (Italy), Amsterdam, Eindhoven, and Utrecht (the Netherlands), Oslo (Norway), Barcelona, Castellon, Santiago, and Valencia (Spain), Stockholm and Umea (Sweden), and London (Great Britain). The repeated measures consist of the within-cities factor "day-of-week" effect (on Wednesday, Thursday , . . . , Tuesday). To obtain some insight of the temporal day-of-week effect, we have taken the average of the standardized usage for each drug and for each day. The results are displayed in Figure 1. We can learn from the patterns that overall drug usage is concentrated around the days that are associated with the weekend, Friday–Sunday. This pattern especially applies to cocaine and ecstasy. Methamphetamine appears to be used most heavily on Monday. These findings are subject to a variety of factors including time lags that are due to the construction of the wastewater measurements. For example, it could take several hours for wastewater to reach its treatment plant depending on the sewerage system and its length.

For each substance (cocaine, ecstasy, cannabis, amphetamine and methamphetamine, all measured in mg per 1000 inhabitants per day) we also investigated the possible effect of the time-independent factor "City size" and "Region" on drug usage. The motivation for the investigation of the "City size" factor is that larger cities tend to have more and larger industries which also dump wastewater into the sewage system. In addition, larger cities attract more commuters and tourists which may be another cause of the effect that sewage water in larger cities is more diluted than in smaller cities.

The estimation of illicit substance usage from sewage biomarkers in larger cities may therefore be less accurate. The time-invariant factor "Region" should capture the spatial (geographic) variation in patterns of illicit drug usage between Western (United Kingdom, the Netherlands, Belgium, France, Spain) and Eastern Europe (Finland, Sweden, Norway, Czech Republic, Italy, Croatia), taking into account the geographical (but not the political divisions)
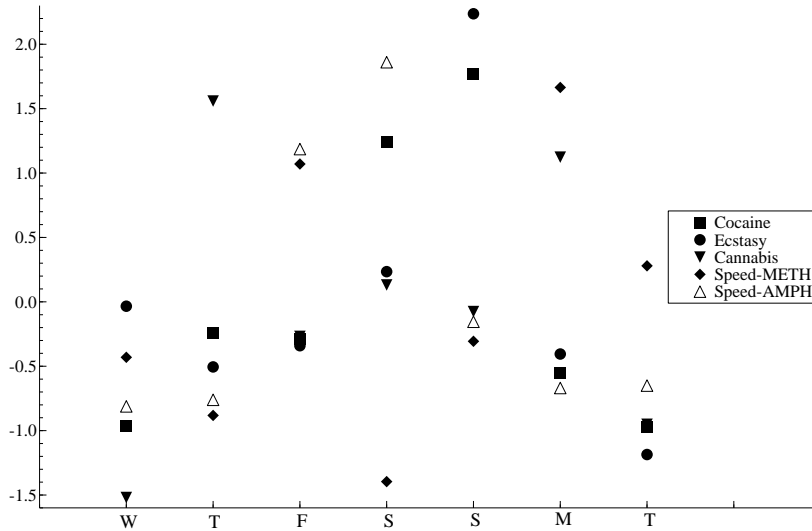
4

Figure 1: Standardized averages of drug usage over 19 European cities, for each day of the week, starting on Wednesday. The x-axis is indexed by day of the week: W for Wednesday,...,T for Tuesday. The y-axis represents average values.

between West and East. In Figure 2 we present the averages of drug use amongst all cities in the West and East parts of Europe. We can clearly observe that the drug use in the West is much higher overall. While cocaine is relatively the most popular drug in both the West and East, amphetamine is also used rather widely in the West. Furthermore we observe that methamphetamine is relatively popular in the East.

The factor "City size" is constructed in the following way. Statistics on city population size and population size of the wastewater treatment plants (WWTP) catchment area were obtained from Ort et al. (2014). These population statistics – including the year to which they apply – are given in Table 1 in increasing order of population size of the wastewater treatment plants (WWTP) catchment area. As a next step, the 19 cities were assigned to the categories "small", "medium", and "large", as shown in Table 1, and the latter categorical variable was used to evaluate the possible effect of city size on drug usage.

For each drug, we have evaluated the effect of the "wealth" of the country. The wealth of a nation is a somewhat subjective notion and can be debated to some lengths. We take the time-independent continuous variable gross domestic product (GDP) per capita of the country in which the city is located as a proxy of the wealth of the city. The GDP per capita data for the year of 2011 has been obtained from Eurostat, and are given in the second column of Table 2.

Finally, we also investigated the possible effect of the continuous time-independent covariate "average number of drug offences per million inhabitants" in each country. The most recent

Table 1: City population and population of the wastewater treatment plants (WWTP) catchment, and its categorization into small, medium, and large cities

| City	WWTP	City Pop	Year	WWTP Pop	Year	Size
Budweis	COV	93,620	(2011)	112,000	(2010)	Small
Umea	Öns	115,473	(2010)	115,800	(2010)	Small
Antwerp	Zuid	498,473	(2011)	117,200	(2010)	Small
Santiago	Silvouta	95,671	(2012)	136,500	(2010)	Small
Castellon	Catellon de la Plana	C: 180,204	(2012)	170,600	(2010)	Small
Turku	Kakola	176,630	(2012)	275,000	(2011)	Small
Utrecht	Utrecht	311,367	(2011)	297,000	(2011)	Small
Stockholm	Henriksdals	C: 847,073	(2010)	315,000	(2009)	Small
	(only 1 of 2 inlets)	M: 1,550,208	(2010)			
Eindhoven	Eindhoven	216,036	(2011)	448,700	(2011)	Medium
Oslo	VEAS	599,230	(2011)	557,000	(2009)	Medium
Zagreb	Central	C: 688,163	Eurostat	650,000	(2001)	Medium
		M: 1,107,623	Eurostat			
Amsterdam	West	779,808	(2011)	694,800	(2011)	Medium
Paris	Seine Centre	C: 2,243,718	(2010)	774,600	(2011)	Medium
		M: 6,507,783	(2006)			
Helsinki	Viikinmäki	M: 1,022,139	(2009)	780,000	(2009)	Medium
Brussels	Noord	1,136,778	(2011)	1,027,300	(2011)	Large
Barcelona	Baix Llobregat	C: 1,620,943	(2012)	1,162,000	(2007)	Large
		M: 3,202,571	(2012)			
Milan	Nosedo	1,295,705	(2009)	1,250,000	(2010)	Large
Valencia	Pinedo I + II & QB	C: 797,028	(2012)	1,839,000	(2011)	Large
	(3 WWTPs)	M: 1,353,250	(2013)			
London	Beckton	8,174,000	(2011)	3,400,000	(2010)	Large

Note: C: City. M: Metropolitan.
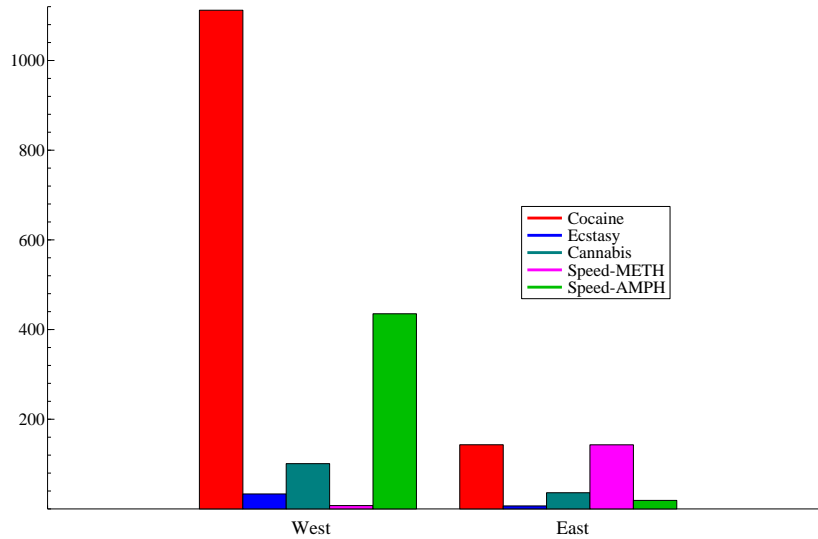
Figure 2: Standardized averages of drug usage over 19 European cities in the West and East parts of Europe.

Table 2: Gross domestic product (GDP), country population, total number of drug offences recorded by the police, and average number of drug offences per million inhabitants

Country	GDP	Population in millions (year)	Number of drug offences	Number of drug offences per million inhabitants
Belgium	29,900	10.6669 (2008)	41,959	3.934
Croatia	15,200	4.2906 (2011)	7,203	1.679
Czech Republic	20,100	10.5622 (2011)	3,834	0.363
Finland	28,800	5.4000 (2012)	20,394	3.777
France	27,200	65.7000 (2012)	190,266	2.896
Italy	25,100	59.4337 (2011)	32,761	0.155
Netherlands	32,900	16.7854 (2013)	16,780	1.000
Norway	46,900	5.0963 (2013)	42,101	8.261
Spain	24,700	46.8159 (2011)	15,220	0.325
Sweden	31,800	9.5964 (2013)	91,997	9.587
Great Britain	27,300	63.1820 (2011)	229,102	3.626

Note: GDP is expressed in Euro, per capita, per year.

7

statistics on the total number of drug offences recorded by the police in each country were obtained from the crime statistics published at www.europeansourcebook.org[2]. The ratio of the total number of drug offences recorded by the police and the country population statistics was then used as a measure of the "number of police recorded drug offences per million inhabitants" in each country. This data is collected in Table 2 and it includes the year for which the total population statistics apply.

3   Results

We have carried out the statistical analyses based on different specifications of the mixed linear models that are reviewed in the Appendix. We have treated the data as repeated measures over seven days of the week, starting on Wednesday, for the 19 cities in our sample. A number of entries are missing. In contrast to a repeated measures ANOVA analysis, linear mixed models are well capable of handling such missing data. We have considered different covariance structures for the disturbances associated with the $n = 19$ cities in our sample. In particular, we have adopted covariance matrices with a compound symmetry (CS) structure, an AR(1) structure, and a Toeplitz (TP) structure for the disturbances of the within-cities factor "Day-of-week". The mean equation of the model includes the between-cities factors "City size" and "Region" together with time-independent covariates "GDP" and "Number of drug offences". This resulting linear mixed model has then be adopted for the analysis of our data set for the five drug substances cocaine, ecstasy, cannabis, methamphetamine and amphetamine, all measured in mg per 1000 inhabitants per day. The estimation results are summarized in terms of their loglikelihood values (scaled by -2) and the Akaike information criterion (AIC) for the three covariance structures in Table 3.

As we have discussed in our treatment of linear mixed models in the Appendix, the CS and AR(1) models are not nested. Hence we need to compare their corresponding AIC values in order to establish which model yields a better fit. However, the TP and CS covariance specifications, on the other hand, are nested. We therefore can compare the fits corresponding to these two specifications by means of the likelihood ratio test. Overall we may conclude that the Toeplitz structure provides the best fit after allowing for the number of parameters. A possible exception is the analysis for cannabis where the more parsimonious AR(1) structure provides the smallest AIC value but the difference with TP is small. A more detailed analysis of the results in Table 3 and the parameter estimation results for each drug are presented and discussed in the remaining part of this section.

---

[2]The European Sourcebook data are kindly provided to us by Paul Smit from the Scientific Research and Documentation Center of the Dutch Ministry of Safety and Justice in the Hague.

Table 3: Linear mixed model goodness-of-fit statistics for all drug substances

Goodness-of-fit statistic	Cocaine	Ecstasy	Cannabis	METH	AMPH
Compound Symmetry, CS: 14 parameters					
−2 loglik	1654.99	766.15	675.48	643.17	1551.61
AIC	1658.99	770.15	679.48	647.17	1555.61
Autoregressive, AR(1): 14 parameters					
−2 loglik	1667.3	757.12	633.41	649.98	1526.61
AIC	1671.3	761.12	637.41	653.98	1530.61
Toeplitz, TP: 19 parameters					
−2 loglik	1627.62	738.02	626.68	630.17	1353.52
AIC	1641.62	752.02	640.68	644.17	1367.52

Note: loglik is the loglikelihood value and AIC is the Akaike information criterion.

3.1 Cocaine

The sewage data entries for cocaine usage in our 19 cities during 7 days are missing for Sunday in Turku, for Monday and Tuesday in Amsterdam, for Wednesday in Utrecht, for Monday in Santiago, and for Saturday in London. Hence we yield a total of 127 ($= 19 \times 7 - 6$) observations.

To select the appropriate covariance structure in the linear mixed model for cocaine, Table 3 reports a better AIC fit for the CS than the AR(1) specifications. The TP and CS covariance structeres are nested and we can compute the likelihood ratio (LR) test statistic as $1654.994 - 1627.620 = 27.374$ for comparing the two specifications. Under the null-hypothesis of no difference in fit, the LR statistic is χ^2-distributed with $19 - 14 = 5$ degrees of freedom. The corresponding p–value is given by $\chi^2_{(0.05;5)}(27.374) = 0.000$ and hence we conclude that the TP model yields the best fit for the cocaine sewage data. This is also confirmed by the AIC values for these two models: the AIC value for the TP model is smallest for cocaine in Table 3.

The F-tests for the fixed main effects of the Toeplitz model are given in the bottom panel of Table 4. We confirm that the omnibus F-test for the Day-of-week factor is highly significant, which implies that some days of the week are different on average from other days of the week. The contrasts for the Day-of-week factor are given in the top panel of Table 4. When using Tuesday as reference category, we observe that the mean cocaine usage on Saturday and Sunday is significantly larger than the usage on Tuesday.

The time-independent factor City size is not significant, nor are the effects of the time-independent covariates Number of drug offences and GDP. On the other hand, the time-independent factor Region is significant, see Table 4. The value of 860.908 for the contrast

Table 4: Estimation results for cocaine usage in 19 European cities
Contrasts for factors "day-of-week", "city size", and "region", and
regression coefficients for covariates "GDP" and "drug offences"

Day-of-week factor	Estimate	Std. Error	df	t	p
Wednesday	-44.406	42.997	10.368	-1.033	0.325
Thursday	38.556	43.714	13.242	0.882	0.393
Friday	32.061	75.193	56.462	0.426	0.671
Saturday	273.054	72.807	44.740	3.750	0.001
Sunday	267.825	71.879	60.156	3.726	0.000
Monday	48.793	61.033	72.413	0.799	0.427
Factor					
City size (small)	-90.971	262.556	12.994	-0.346	0.735
City size (medium)	-115.962	287.487	12.984	-0.403	0.693
Region (West)	860.908	262.849	12.994	3.275	0.006
Covariate					
GDP	0.018	0.023	12.979	0.789	0.444
Drug offences	-0.041	0.052	12.987	-0.791	0.443

Fixed main effects of factors and covariates

Source	F	$df1$	$df2$	p
Day-of-week	5.023	6	37.142	0.001
City size	0.090	2	12.990	0.914
Region	10.728	1	12.994	0.006
GDP	0.622	1	12.979	0.444
Drug offences	0.625	1	12.987	0.443

Note: TP covariance structure with 19 estimated parameters.

Estimated marginal means of reference categories: Tuesday = 555.645; City size (Large) = 712.606; Region (East) = 213.174.

for Region implies that – keeping all other independent variables constant – the usage of cocaine in the Western European countries is about 861 mg per 1000 people per day higher than that in the Eastern European countries, taking into account the geographical and not the political division between West and Eastern Europe. In particular, the highest values (>1000 mg/1000 inhabitants/day) were observed in the United Kingdom, the Netherlands, Belgium and Spain, whereas the lowest values (<200 mg/1000 inhabitants/day) were observed in Croatia, Czech Republic, Norway, Sweden and Finland. A clear geographical difference in cocaine usage between Western and Eastern Europe, with higher use in Western Europe, has also been found by Ort et al. (2014).

Adding a first-order interaction effect for Day-of-week by City size and Day-of-week by Region to the previous Toeplitz covariance main effects model yields a value of 1394.212 for -2 times the loglikelihood with 30 parameters. Comparing these two nested models with the likelihood ratio test, we obtain $1627.620 - 1394.212 = 233.408$. Since – under the null-hypothesis of no difference in fit – the LR statistic is χ^2 distributed with $37 - 19 = 18$ degrees of freedom and with its p–value given by $\chi^2_{(0.05;18)}(233.408) = 0.000$. We can conclude that the TP model with interactions yields a significantly better fit than the TP model without interactions. This is also confirmed by the value of the AIC which is now 1408.212, a much smaller value than for the main effects model.

The results of this final analysis are given in Table 5. It becomes clear from these results that the interaction for Day-of-week by City size is not significant. On the other hand, the interaction of Day-of-week by Region is significant. This is mostly due to the fact that the difference between cocaine usage on Saturday and Sunday and the rest of the week is much more pronounced in the Western than in the Eastern European countries, see Figure 3.

Table 5: Fixed main and interaction effects of day-of-week, city size, region, GDP, and drug offences on cocaine usage in 19 European cities

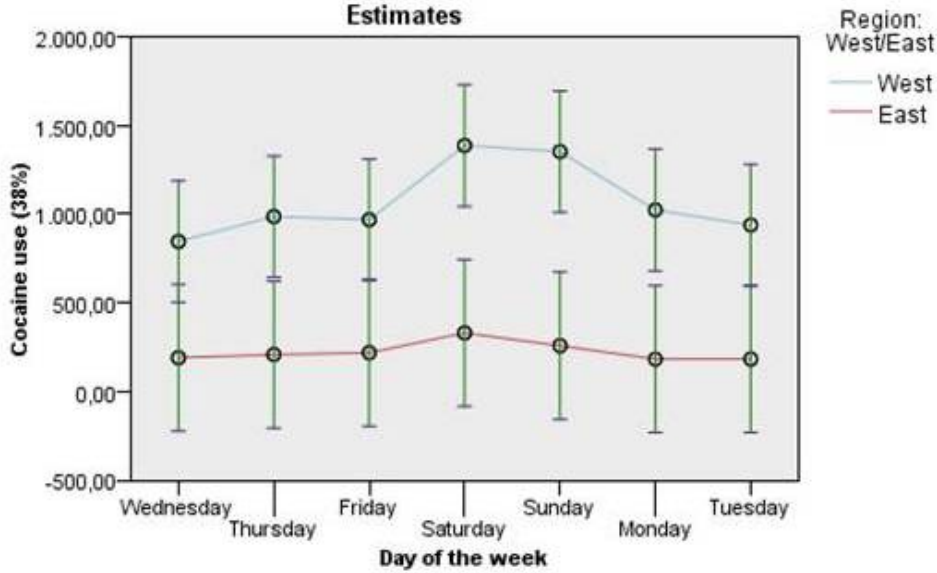| Source	F	$df1$	$df2$	p
Day-of-week	9.060	6	19.857	0.000
City size	0.082	2	13.049	0.922
Region	10.145	1	13.030	0.007
GDP	0.594	1	12.985	0.455
Drug offences	0.554	1	12.990	0.470
Day*City size	1.379	12	20.735	0.252
Day*Region	3.961	6	20.705	0.009

Figure 3: Interaction between day-of-week and region for cocaine usage in 19 European cities.

3.2 Ecstasy

We have no sewage data entries for ecstasy usage in the cities of Oslo, Castellon, Stockholm, and Umea. For the remaining 15 cities, data for ecstasy usage are missing on Sunday in Turku, on Monday and Tuesday in Amsterdam, on Wednesday in Utrecht, on Monday in Santiago, and on Saturday in London. Moreover, for its six non-missing observations we have encountered that the city of Utrecht is an outlier that completely dominates the analysis. As mentioned in Thomas et al. (2012), the extremely high scores on ecstasy usage in this city are related to the dismantling of an illegal ecstasy production facility at the time the sewage samples were taken. After removing Utrecht from the analysis, we have analyzed the remaining 93 observations and have obtained the results presented in Table 3.

In terms of the AIC values for the non-nested CS and AR(1) models, we can conclude that the AR(1) covariance structure is preferred. When we consider the LR test statistic for the nested TP and AR(1) covariance models, we have $757.123 - 738.024(TP) = 19.099$ which, under the null-hypothesis of no difference, is χ^2-distributed with $19 - 14 = 5$ degrees of freedom with its p–value given by $\chi^2_{(0.05;5)}(19.099) = 0.002$. We can conclude that the TP covariance structure yields the best fitting model. This is also confirmed by the AIC values for the models, see Table 3.

The estimated fixed effects for this model are presented in the bottom panel of Table 6. The very significant effect for the factor Day-of-week is prominent. The estimated contrasts for the factor Day-of-week are given in the top panel of Table 6. When using Tuesday as reference category, the significant omnibus F-test for Day-of-week is probably due to the fact that the

ecstasy mean estimate for Sunday is significantly higher than the mean esimates for the other days of the week; see also Figure 1.

Table 6: Estimation results for ecstasy usage in 14 European cities
Contrasts for factors "Day-of-week", "City size", and "Region", and regression coefficients for covariates GDP and "Number of drug offences per million inhabitants"

Day-of-week	Estimate	Std. Error	df	t	p
Wednesday	7.062	10.788	37.968	0.655	0.517
Thursday	3.014	9.610	38.081	0.314	0.756
Friday	4.427	6.832	24.930	0.648	0.523
Saturday	10.400	7.406	56.422	1.404	0.166
Sunday	24.852	6.870	43.518	3.617	0.001
Monday	6.797	5.711	47.112	1.190	0.240
Factor					
City size (Small)	5.328	6.492	6.649	0.821	0.440
City size (Medium)	24.352	6.330	6.766	3.847	0.007
Region (West)	9.026	6.511	7.074	1.386	0.208
Covariate					
GDP	0.002	0.001	7.146	2.588	0.035
Drug offences	-0.007	0.002	6.909	-3.493	0.010

Fixed main effects of factors and covariates

Source	F	$df1$	$df2$	p
Day-of-week	5.342	6	26.625	0.001
City size	8.028	2	6.882	0.016
Region	1.922	1	7.074	0.208
GDP	6.698	1	7.146	0.035
Drug offences	12.201	1	6.909	0.010

Note: TP covariance structure with 19 estimated parameters.

Estimated marginal means of reference categories: Tuesday = 14.484; City size (Large) = 12.670; Region (East) = 18.050.

The main effect of City size on ecstasy drug usage is significant, whereas the effect of Region is not, see Table 6. The significant effect of City size may be due to the usage of ecstasy being 24.35 mg per 1000 inhabitants per day higher for medium-sized cities compared to the usage of ecstasy in large cities (the reference category). The effect of covariate GDP on ecstasy usage is also significant. Since the estimated regression coefficient for GDP equals the positive value of

0.002 and is significant, the usage of ecstasy increases when the wealth of a country increases. In particular, when keeping all other variables constant, an increase of 100 euro in the GDP per capita leads to an increase in usage of ecstasy by 0.2 mg per 1000 inhabitants per day. The effect related to the number of drug offences is negative and significant. The estimated regression coefficient of -0.007 implies that an increase of 100 police recorded drug offences per million inhabitants in a country is associated with a *decrease* in ecstasy usage of 0.7 mg per 1000 inhabitants per day. The causality related to this effect can be viewed in different ways. When more drug related offences are treated by the police, the opportunities for ecstasy usage become less due to supply problems. On the other hand, when drug related offences are not treated very heavily by the police, it may reflect a more liberal attitude to drug usage. Hence drugs become socially more acceptable and its usage more widespread.

Table 7: Fixed main and interaction effects of Day-of-week, City size, Region, GDP, and Drug offences on ecstasy usage in 14 European cities

Source	F	$df1$	$df2$	p
Day-of-week	3.172	6	19.015	0.025
City size	1.384	2	16.559	0.278
Region	3.044	1	17.819	0.098
GDP	6.468	1	6.461	0.041
Drug offences	11.775	1	6.218	0.013
Day*City size	1.211	12	18.755	0.344
Day*Region	0.651	6	18.441	0.689

To verify whether the interactions for Day-of-week versus City size and Day-of-week versus Region have a role to play in our analysis of excstasy usage, we have estimated the parameters of the same model but with these interactions included as well. The results are presented in Table 7. From the relatively high p-values, we can conclude that both interactions do not play a significant role in our analysis of ecstasy usage.

3.3 Cannabis

The cities of Antwerp, Brussels, Helsinki, Turku, Oslo, Stockholm, Umea, and London have not provided the sewage entries for cannabis usage. For the remaining 11 cities, data for cannabis usage are also missing for Monday and Tuesday in Amsterdam, for Wednesday in Utrecht, and for Monday in Santiago. Hence we only have 73 ($= 11 \times 7 - 4$) observations available for our analysis of cannabis usage. We have estimated the parameters in the model specifications with CS, AR(1) and TP covariance structures. The goodness-of-fit results are presented in Table 3. Comparing the AIC criteria of the non-nested AR(1) and CS covariance models, we can conclude

that the AR(1) model yields the smallest AIC. The difference in (scaled) loglikelihood values between the nested AR(1) and TP models is $633.414 - 626.681 = 6.733$ which, under the null-hypothesis of no difference in fit, is χ^2-distributed with $19 - 14 = 5$ degrees of freedom. The corresponding p–value is $\chi^2_{(0.05;5)}(6.733) = 0.241$ and we therefore conclude that the AR(1) covariance structure yields the best fitting model. This is confirmed by the AIC values for the two models which is smaller for the AR(1) model as well.

The estimation results for the fixed main effects are presented in Table 8. None of the main effects for Day-of-week, City size, Region, GDP, and Number of drug offences are estimated as significant, irrespective of the covariance structure used, that is CS, AR(1) or TP. We suspect this result is due to the relatively small sample size of 73 observations from which 14 parameters need to be estimated. For our preferred covariance structure, this is the AR(1) specification, the main estimation results are given in Table 8. We do not present the estimations results for the models with the interaction effects for Day-of-week by City size and Day-of-week by Region included. These results would have been based on less than 5 observations per estimated parameter and we have concluded that this is too small.

Table 8: Fixed main effects of Day-of-week, City size, Region, GDP and Drug offences per million inhabitants on cannabis usage in 11 European cities

Source	F	$df1$	$df2$	p
Day-of-week	1.767	6	52.405	0.124
City size	0.495	2	5.591	0.634
Region	0.141	1	5.470	0.721
GDP	1.483	1	5.577	0.272
Drug offences	0.004	1	5.587	0.949

3.4   Speed drugs : methamphetamine and amphetamine

The data for methamphetamine (METH) usage – as measured in mg/day per 1000 inhabitants – are not available for the cities of Brussels, Zagreb, Turku, Paris, Amsterdam, Eindhoven, Utrecht, and Castellon. For the remaining 11 cities, we have missing entries for Monday in Santiago and for Saturday in London. The number of observations $75 = 11 \times 7 - 2$ is also small in this case. The best fitting covariance structure which is the Toeplitz model as we have concluded from the general discussion for Table 3. The corresponding estimation results are presented in Table 9. We find that the estimated factors Day-of-week and Drug offences are not significant. However, the effect of City size is highly significant. This result may be due to the fact that METH usage in medium-sized cities is significantly larger than that in large cities which is the reference category. The effects of Region and GDP on METH usage are

only marginally significant. There is some indication that METH usage is smaller in West than in East European countries while higher METH usage is associated with lower GDP. In the analysis of Ort et al. (2014), differences in METH usage in Eastern and Western Germany are found.

Table 9: Estimation results for METH usage in 11 European cities
Contrasts for factors "Day-of-week", "City size", and "Region", and regression coefficients for covariates GDP and "Number of drug offences per million inhabitants"

Day-of-week	Estimate	Std. Error	df	t	p
Wednesday	-4.181	3.861	6.738	-1.083	0.316
Thursday	-6.839	9.917	30.141	-0.690	0.496
Friday	4.654	10.113	30.742	0.460	0.649
Saturday	-17.635	10.600	28.812	-1.664	0.107
Sunday	-3.449	10.162	37.752	-0.339	0.736
Monday	0.723	8.998	45.372	0.080	0.936
Factor					
City size (Small)	28.024	23.843	5.005	1.175	0.293
City size (Medium)	353.639	40.812	4.937	8.665	0.000
Region (West)	-60.621	24.818	4.960	-2.443	0.059
Covariate					
GDP	-0.007	0.003	4.946	-2.344	0.067
Drug offences	-0.004	0.005	4.970	-0.656	0.541

Fixed main effects of factors and covariates

Source	F	$df1$	$df2$	p
Day-of-week	1.341	6	17.580	0.291
City size	38.769	2	4.970	0.001
Region	5.967	1	4.960	0.059
GDP	5.493	1	4.946	0.067
Drug offences	0.430	1	4.970	0.541

Note: TP covariance structure with 19 estimated parameters.

Estimated marginal means of reference categories: Tuesday = 131.757; City size (Large) = 0.718; Region (East) = 158.250.

When we account for the interaction effects between Day-of-week and City size and Day-of-week and Region, we find significant interaction effects for the Day-of-week and City size. However, due to the small sample size relative to the large number of estimated parameters,

that is 75 observations for 37 parameters, these findings may not be very stable. Hence we do not consider these results further. For the usage of amphetamine (AMPH), we have obtained very similar results as those obtained for cannabis and therefore we do not discuss these results here.

4   Discussion

For our data set of sewage data from 19 European cities, the average amount of cocaine found in the sewage samples is significantly larger on Saturday and Sunday when compared to other days of the week. For the 14 cities from which ecstasy measurements were taken of the sewage samples, the average amount of ecstasy is significantly larger on Sunday than on any other day of the week. For the other drugs, no significant differences in usage are found between the days of the week. Our analysis provides a solid statistical basis to confirm this conclusion that was also found by Thomas et al. (2012): cocaine and ecstasy are typically recreational drugs.

An effect of city size, measured by the size of the population of the wastewater treatment plants catchment, is found in the usage of ecstasy and methamphetamine. The medium-sized cities of Eindhoven, Olso, Zagreb, Amsterdam, Paris, and Helsinki have significantly higher usages of these two drugs than the large cities of Brussels, Barcelona, Milan, Valencia, and London.

Significant regional (spatial) effects are found in the usage of cocaine. The usage of cocaine in the Western European countries is about 861 mg per 1000 people per day larger than the usage of cocaine in the Eastern European countries, taking into account the geographical and not political division between West and East. In particular, the highest values (>1000 mg/1000 people/day) were observed in the United Kingdom, the Netherlands, Belgium and Spain, whereas the lowest values (<200 mg/1000 people/day) were observed in Croatia, Czech Republic, Norway, Sweden and Finland. A clear geographical difference in cocaine usage between Western and Eastern Europe, with higher usage in Western Europe, has also been found by Ort et al. (2014).

With respect to the effect of the economic situation on drug usage in a country, a positive relation is found between GDP and the usage of ecstasy, in the sense that countries with a higher GDP are associated with the usage of larger amounts of ecstasy. On the other hand, a negative relation is found between GDP and the usage of methamphetamine, in the sense that countries with a higher GDP are associated with the usage of lower amounts of methamphetamine. For methamphetamine this relation is only indicative ($p = 0.067$), while it is significant for ecstasy ($p = 0.035$). For cocaine, cannabis and amphetamine such relation is not found.

We have observed a significant negative relationship between the per capita number of drug offences registered by the police in a country and the usage of ecstasy, in the sense that higher

numbers of per capita police-recorded drug offences in a country are associated with the usage of smaller amounts of ecstasy. This seems to imply that the more successful the police are in the apprehension of citizens involved in drug related criminal offences in a country the smaller the usage of illegal drugs is, at least as far as ecstasy is concerned. For cocaine, methamphetamine, amphetamine, and cannabis we have not found such relation.

References

European Monitoring Centre for Drugs and Drug Addiction (2011). The state of drugs problem in Europe. Luxembourg: Office for Official Publications of the European Communities.

Ort, C., A. L. N. van Nuijs, J.-D. Berset, L. Bijlsma, S. Castiglioni, A. Covaci, P. de Voogt, E. Emke, D. Fatta-Kassinos, P. Griffiths, F. Hernández, I. González-Marino, R. Grabic, B. Kasprzyk-Hordern, N. Mastroianni, A. Meierjohann, T. Nefau, M. Ostman, Y. Pico, I. Racamonde, M. Reid, J. Slobodnik, S. Terzic, N. Thomaidis, and K. V. Thomas (2014). Spatial differences and temporal changes in illicit drug use in Europe quantified by wastewater analysis. *Addicition 109*, 1338–1352.

Stroup, W. W. (2013). *Generalized Linear Mixed Models: Modern Concepts, Methods and Applications.* New York: CRC Press.

Thomas, K. V., L. Bijlsma, S. Castiglioni, A. Covaci, E. Emke, R. Grabic, F. Hernández, S. Karolak, B. Kasprzyk-Hordern, R. H. Lindberg, M. L. de Alda, A. Meierjohann, C. Ort, Y. Pico, J. B. Quintana, M. Reid, J. Rieckermann, S. Terzic, A. L. N. van Nuijs, and P. de Voogt (2012). Comparing illicit drug use in 19 European cities through sewage analysis. *Science of the Total Environment* (432), 432–439.

Twisk, J. R. (2013). *Applied Longitudinal Data Analysis for Epidemiology: A Practical Guide* (2nd ed.). Cambridge: Cambridge University Press.

Appendices

A   Methodology

Linear mixed models can be best conceived of as a generalization of repeated measures analysis of variance (ANOVA) that allows for the analysis of an identical number of repeated observations made on several subjects. The repeated measures ANOVA itself can be rather restrictive as it cannot cope with missing observations. As soon as one observation for a subject is missing,

the subject must be completely dropped from the analysis. In this way, we need to remove data unneccessarily. This is a rather inefficient approach, especially as so much time and (financial) effort is given to collect the data set.

Another important restriction of repeated measures ANOVA can be described as follows. The treatment of the time dependencies in repeated measures ANOVA can only be handled in the following two ways. The first option is to assume a completely unrestricted covariance structure for the residuals of the repeated measures. To discuss this option in more detail, let n denote the number of repeated measures in the analysis. Consider the case of $n = 4$, for example, the unrestricted covariance matrix can be written as

$$\Sigma_{\text{UN}} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} & \sigma_{24} \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 & \sigma_{34} \\ \sigma_{14} & \sigma_{24} & \sigma_{34} & \sigma_4^2 \end{bmatrix}. \tag{1}$$

In this approach, all observations at time i $(i = 1, \ldots, n)$ have a unique variance and observations within the same subject at every pair of times have a unique covariance. This approach requires the estimation of $n(n+1)/2$ covariance parameters, that is 10 parameters when $n = 4$. However, the statistical tests based on this approach also have a tendency to lack power. It implies that null-hypotheses are often not rejected even if they are false.

The second option is to assume a highly restrictive compound symmetry covariance structure for the residuals of the repeated measures. When $n = 4$, the compound symmetry covariance matrix can be written as

$$\Sigma_{\text{CS}} = \sigma^2 \begin{bmatrix} 1 & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho \\ \rho & \rho & 1 & \rho \\ \rho & \rho & \rho & 1 \end{bmatrix}. \tag{2}$$

In this case, we assume that the variances for observations at every time i are all equal while the observations within the same subject at every pair of times have the same covariance, irrespective of the length of the time interval between the two measurements. This approach only requires the estimation of two covariance parameters, irrespective of the value of n. However, the imposed restrictions are rather unrealistic, and – if not satisfied – results in standard errors that are too small, and therefore in F- and t-values that are too large. This approach may therefore result in type I errors: null-hypotheses are rejected even if they are actually true.

When we adopt linear mixed models, subjects with missing observations do not need to be removed from the analysis while other covariance structures can still be applied and tested including the two aforementioned unrestricted and compound symmetry covariance structures. In particular, we further consider two other covariance structure models. The first specification

is for an autoregressive (of order 1) covariance structure which can be written as

$$\Sigma_{\text{AR}(1)} = \sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}, \tag{3}$$

where we typically assume the stationary condition of $|\rho| < 1$. This structure is closely related to the compound structure in (2) but the impact of the covariance vanishes exponentially when measurement entries are taken further apart from each other. In this covariance structure, correlations between observations of d measurements apart are all equal to ρ^d. Correlations are therefore assumed to become smaller and smaller as observations are further separated in time. The AR(1) covariance structure (3) requires the estimation of only two unknown parameters, irrespective of the value of n.

The second covariance structure under investigation is the Toeplitz specification. The Toeplitz covariance structure can be written as

$$\Sigma_{\text{TP}} = \sigma^2 \begin{bmatrix} 1 & \rho_1 & \rho_2 & \rho_3 \\ \rho_1 & 1 & \rho_1 & \rho_2 \\ \rho_2 & \rho_1 & 1 & \rho_1 \\ \rho_3 & \rho_2 & \rho_1 & 1 \end{bmatrix}, \tag{4}$$

where the correlations between observations of d measurements apart are all assumed to be equal to ρ_d. In the Toeplitz covariance structure, the correlations ρ_d are allowed to take on different values; they are not restricted to a slowly decaying process. The Toeplitz covariance structure requires the estimation of n parameters, that is the number of repeated measures in the data.

In all analyses we are comparing the results obtained with the compound symmetry model (2), the AR(1) model (3), and the Toeplitz model (4). To compare their performance we use the likelihood ratio test if the models are nested, and the Akaike information criterion (AIC) otherwise. The compound symmetry model (2) is not nested within AR(1) model (3) because no restriction on the AR(1) covariance structure (3) will ever reduce it to the compound symmetry covariance structure (2). The compound symmetry model (2) is nested within the Toeplitz model (4), on the other hand, because the restriction $\rho_i = \rho$ for $i = 1, \ldots, n$ in the Toeplitz covariance structure (4) automatically reduces it to the compound symmetry covariance structure (2). The AR(1) model (3) is also nested within the Toeplitz model (4) because the restriction $\rho_i = \rho_1^i$ for $i = 2, \ldots, n$ in the Toeplitz covariance structure (4) automatically reduces it to the AR(1) covariance structure (3). In our study we do not consider results for the unrestricted model (1) because it requires the estimation of too many parameters. In this

setting the number of parameters can exceed half of the total number of observations. This will result in various sorts of convergence issues in its implementation. In all analyses we used the so-called Satterthwaite approximation to the degrees of freedom. It implies that the degrees of freedom can vary across tests. This can be useful for smaller sample sizes, unbalanced data or more complicated covariance structures.

For further details on linear mixed models we refer to Twisk (2013) and Stroup (2013).