European Monitoring Centre
for Drugs and Drug Addiction

INSIGHTS

MONITORING DRUG USE IN THE DIGITAL AGE: STUDIES IN WEB SURVEYS

Comparing two approaches for estimating cannabis consumption: raking and regression

Jonathan P. Caulkins,[a] Kyle Furlong,[a] Brandon McComas,[a] Martha Nogueira Domingues,[a] Aatir Siddique[a] and Beau Kilmer[b]

[a] Carnegie Mellon University, Heinz College of Information Systems and Public Policy

[b] The RAND Corporation, Drug Policy Research Center

Abstract: Knowing the size of illicit drug markets is important for understanding their revenues and making informed policy decisions. While general population surveys (GPS) can provide an estimate of the number of potential users of such markets, many GPS are limited by only collecting information on the prevalence and frequency of drug use, missing detailed data on the quantity and value of drugs consumed. In contrast, web surveys are much cheaper to implement and are often constructed with a more specific focus. This chapter explores two statistical methods of bringing together the complementary strengths of GPS and web surveys to develop improved estimates of overall quantities of drugs consumed — namely, raking and regression. By doing such, it shows how the rich information collected by online surveys could potentially be extrapolated to larger populations. This chapter applies these two methods to data captured by the European Web Survey on Drugs (EWSD) and six European countries' respective GPS to measure total cannabis consumption in the general population. While both of these methods have their own limitations, this chapter shows that they could still represent an improvement over making projections based on simple rules of thumb or having no estimates at all.

Introduction

Knowing the size of illicit drug markets is important for understanding their revenues and making informed policy decisions (Kilmer et al., 2011). General population surveys (GPS) can provide an estimate of the number of potential users of such markets, as these surveys recruit a representative sample to estimate the overall prevalence of drug use. While some people are excluded from the sampling frame of GPS and respondents may misreport their use, these surveys are generally thought to perform reasonably well in terms of measuring the number of users of more common and less stigmatised drugs such as cannabis (e.g., Kilmer et al., 2014; Rhodes et al., 2012).

However, when trying to estimate the size of drug markets (e.g. in terms of overall quantities of drugs consumed or such markets' monetary value), many GPS are limited by only collecting information on the prevalence and frequency of drug use, missing detailed data on the quantity and value of drugs consumed. In contrast, web surveys are much cheaper to implement and are often constructed with a more specific focus — for example, surveys that concentrate just on substance use. Of particular interest here is that some web surveys obtain granular data from people who use drugs about their consumption patterns. Such detailed data can then fill the gaps in GPS and help develop improved estimates of the size of drug markets in terms of the overall quantities of drugs consumed. This can in turn provide an indication of the estimated value of such markets if data also exists on drug prices.

While the limitations of web surveys have been covered elsewhere in this publication, it is important to mention those of relevance to the topic of this report. In particular, certain individuals may be under-represented in web surveys — such as those with limited access to the internet — while others may be more likely to complete the survey. For example, in the first wave (2016) of the European Web Survey on Drugs (EWSD), high-frequency cannabis users were greatly over-represented. However, this problem also represents an opportunity, since high-frequency substance users account for the great bulk of consumption and are rare among GPS respondents. If, for example, 10 % of a country's population uses cannabis and 20 % of this group use daily or near daily, then as a result only 2 % of GPS respondents will be high-frequency users (if the GPS reaches cannabis users in a representative way and if such users accurately report their use). Hence, both web surveys and GPS have complementary strengths and weaknesses. GPS produce estimates that are representative of the population surveyed, but only for prevalence and not for consumption levels. Web surveys provide rich information about consumption, including by all-important heavy users, but not in a way that can be extrapolated to any larger population beyond the respondents.

This chapter explores two methods of bringing together the complementary strengths of GPS and web surveys to develop improved estimates of overall quantities of drugs consumed — namely, raking and regression. Raking re-weights responses to a web survey in an attempt to make the weighted averages or totals match those of the general population. Regression is used to impute values for what GPS respondents might have said if they had been asked about the frequency of their drug use and the quantities consumed. This chapter applies these two methods to data captured by the EWSD and the countries' respective GPS to measure total cannabis consumption in the general population. This is pertinent to the literature described in the EMCDDA (2012) study on the cannabis market in Europe and by van Laar et al. (2013), among others.

Methods

Data sources

We use data from the first wave of the European Web Survey on Drugs (EWSD) from 2016, which included six countries: Croatia, Czechia, France, the Netherlands, Switzerland and the United Kingdom, in addition to the respective countries' GPS for past-month users as reported in the EMCDDA's Statistical Bulletin for 2017. Table 1 shows the EWSD sample sizes and missing data information by country. Many respondents did not fill out questions regarding age or gender, which precludes using those respondents in connection with the GPS.

Method 1: raking

When a survey sample covers segments of the target population 'in proportions that do not match the proportions of those segments in the population itself', raking (also known as 'raking ratio estimation' or 'sample-balancing') can be employed to 'improve the relation between the sample and the population' (Battaglia et al., 2009). Raking is done by adjusting the sampling weights of the cases in a survey sample 'so that the marginal totals of the adjusted weights on specified characteristics ... agree with the corresponding totals for the population' (Battaglia et al., 2009).

In this chapter, raking is used in an attempt to calculate weights for respondents to the EWSD that allow weighted counts to match the characteristics of the general population as reflected in the GPS. For example, if 70 % of the EWSD respondents' ages fell within a certain range but in the GPS only 35 % of past-month users were in that age bracket, raking would weight each of the EWSD respondents by 0.5 because 70 % × 0.5 = 35 %. Opinions vary on how successful re-weighting is likely to be. Some studies report success in eliminating bias (e.g. Dever et al., 2008); others caution that weighting adjustments may have modest effects and may not eliminate differences in terms of what a GPS would find if it asked that question (e.g. Loosveldt and Sonck, 2008; Yeager et al., 2011).

TABLE 1
Extent of missing data

	Initial respondents	Number who reported age	Number who reported gender	Final respondents
Croatia	9 142	4 412	4 626	4 410
Czechia	1 058	565	580	565
France	4 849	2 514	2 605	2 512
Netherlands	1 238	755	758	755
Switzerland	2 862	1 543	1547	1 542
United Kingdom	1 216	364	438	358
Total	**20 365**	**10 153**	**10 554**	**10 142**

Method 2: regression

While raking is applied to re-weight EWSD respondents' answers; regression uses EWSD responses to impute grams per day of use for survey respondents who were not directly asked for that information in the GPS. We calculated the daily average consumption for each user in the EWSD sample and regressed it on categorical variables, coding for all available variables that tend to be correlated with consumption — namely, gender, age, frequency, education and income:

$$daily\ grams_i = \beta_1 gender_i + \beta_2 age_i + \beta_3 frequency_i + \beta_4\ income_i + \beta_5 education_i + \varepsilon_i$$

Combining information about herbal and resin consumption

The EWSD asks past-month users about their consumption of two types of cannabis: resin and herbal. Our aim is to establish a single aggregate number for overall cannabis consumption. However, it would be misleading to simply add kilograms of resin to kilograms of herbal because resin tends to be about twice as potent. Thus, the total aggregate number for overall cannabis consumption should be understood as the weight of herbal cannabis that would supply roughly the same amount of THC as the combination of herbal and resin actually consumed. This is inspired by the idea of a 'standard drink' in alcohol, which combines beer, wine and spirits volumes in a manner that adjusts for their varying alcohol content. The 2:1 ratio is a round number approximation taken from data on the potency of resin and herbal cannabis gathered in France, Croatia and Czechia for 2013–2015 and obtained from the EMCDDA Statistical Bulletin (EMCDDA, 2017b; data for Switzerland, the Netherlands and the United Kingdom were not available). Furthermore, an analysis conducted in France in mid-2016 found the THC content in resin and herbal to be about 23 % and 13 % respectively, which is also roughly consistent with a 2:1 ratio (Dujourdy and Besacier, 2017).

Further, in the 2016 wave of the EWSD it was found that most respondents (4 207 out of 6 537) consumed only one type of cannabis, but a sizable minority reported using both resin and herbal in the same month. The survey included separate questions about days of use of resin and herbal respectively, but did not ask about days of use of cannabis generally. This makes it ambiguous whether someone who answers '15' to both of those questions used cannabis on just 15 days, that is, they consumed both herbal and resin on each of those 15 days, or whether that person used cannabis on as many as 30 days in the past month, if never using both forms of cannabis on the same day. This ambiguity turns out to be challenging but not problematic, as shown below.

TABLE 2

User classification based on highest and capped sum metrics

Days used cannabis in past month, highest metric	Days used cannabis in past month, capped sum metric					
	0	1–3	4–9	10–19	20+	Total
0	–	–	–	–	–	–
1–3	–	1 100	108	–	–	1 208
4–9	–	–	792	103	–	895
10–19	–	–	–	932	227	1 159
20+	–	–	–	–	3 276	3 276
Total	**–**	**1 100**	**900**	**1 035**	**3 503**	**6 537**

Given this indeterminacy, we proceed in two ways that bound the range of possibilities. These provide similar results for aggregate days of use, so we simply report the average obtained using the two approaches. The first method assumes that consumption of herbal and resin takes place on the same day, so it simply takes the highest frequency of days of use for resin or herbal. The second method adds the days of resin and herbal use, capping the sum at 30 since there are generally only 30 days in a month. A cross-tab of user classification (Table 2) using these two metrics shows that users are categorised into the same bin most of the time. For statistics about frequency, we use the first metric. When estimating quantities consumed, we use both metrics and average the two results.

Estimating monthly and daily consumption

The 2016 EWSD allowed users to describe amounts used (in grams) per day of use in a structured way, guided by images of piles of cannabis juxtaposed with common objects such as a coin or credit card to help users self-report quantities more accurately. To estimate monthly consumption, we calculate the monthly estimates for both resin and herbal and combine them as follows:

$$(1) \qquad (Q_{Hi}F_{Hi}) + (Q_{Ri}F_{Ri})$$

Q_{Hi} Average quantity of herbal cannabis user 'i' consumed per day of use in past month (calculated)

F_{Hi} Frequency in number of days used herbal cannabis in last 30 days (given by the user)

Q_{Ri} Average quantity of cannabis resin user 'i' consumed per day of use in past month (calculated)

F_{Ri} Frequency in number of days used resin cannabis in last 30 days (given by the user)

Consumption per day of use equals monthly cannabis consumption divided by the number of days of cannabis use in the past month. Since there are two ways to estimate the number of days of cannabis use, this ratio was computed in two ways, and the final value was the average of those two ratios. In particular, consumption per day of use was computed both by dividing monthly consumption by both the capped sum of days of use and also by the greater of days of resin use and days of herbal use. We then took the average of those two ratios. The results are similar with or without the 274 respondents who reported consumption for both types of cannabis but not age or gender.

Additionally, as highlighted previously, we multiply the grams of resin consumed by two to account for the fact that resin is generally of higher potency. In this way, monthly and daily consumption weights reflect the amount of herbal that would deliver the same amount of THC as the actual mix of herbal and resin consumed.

Results

Method 1: raking

We apply raking to match the distributions of age, gender and past-month frequency of use within each country as closely as possible to those of the respective countries' GPS results for past-month users as reported in the EMCDDA Statistical Bulletin (EMCDDA, 2017a). There is, however, a mismatch in the ages covered by the two data sources. While the EWSD only includes people aged 18 or over, the countries' respective GPS incorporates 15–17-year-olds as well. As such, the youngest age range for both sources is referred to as up to 24-year-olds.

Table 3 shows the need for such re-weighting. EWSD respondents tend to be younger and much more likely to report using cannabis 20 or more times in the past month than GPS respondents.

One limitation is that raking can only include the EWSD respondents who gave their age, gender, frequency of past month use and past month consumption, and many refused to report their age or gender. If these respondents typically consume more or less than those who provided age and gender data, controlling for other observables, then that non-response could bias results.

Another limitation is that raking assigns some individuals a large weight, as shown in Figure 1. Large weights, shown as outliers in this box and whiskers plot (Figure 1), make estimates more variable. If an individual who is assigned a high weight gives an atypical answer, then the distorting effects of that answer are multiplied.

A final and fundamental limitation is the implicit assumption that after controlling for the variables used in the raking, the EWSD respondents' answers are similar to those of the general population. This is plausible given that frequency of use is so strongly related to quantity consumed, but it need not be the case. For example, EWSD respondents might be wealthier than non-respondents, and those with more money might buy more cannabis, but the raking did not control for income.

Method 2: regression

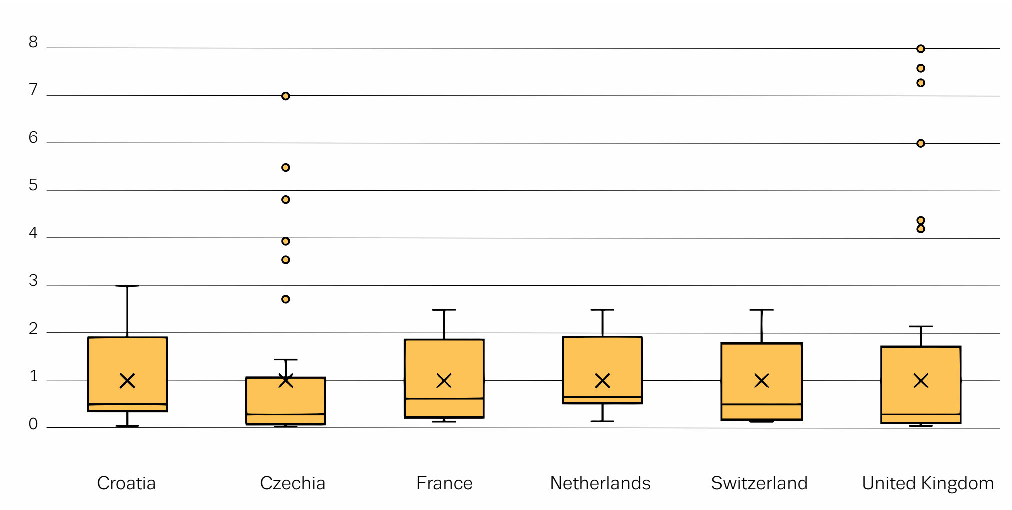
People who report using on many days also describe using more grams per day of use, although the patterns vary by gender and age. For example, men tend to consume more grams per day of use than women for all frequencies of use, but the gap is larger among infrequent users. Likewise, young people in the EWSD sample tend to consume more per day of use than older users across all the frequencies of past-month use, with the difference perhaps most pronounced among heavy users. Among those using on 20+ days per month, those in the youngest age category report consuming 1.31 grams

TABLE 3
Quantity consumption distribution

	Number of observations with non-missing values	Proportion males (%)		Proportion up to age 24 (%)		Proportion using on 20+ days per month (%)	
		EWSD	GPS	EWSD	GPS	EWSD	GPS
Croatia	815	66	76	49	42	55	16
Czechia	107	83	81	28	45	56	8
France	1 421	63	70	64	39	58	33
Netherlands	337	63	73	51	42	37	28
Switzerland	874	71	72	63	42	51	26
United Kingdom	180	71	73	53	32	38	8

The EWSD covers people aged 18 or above. Questions on frequency of use may differ in general population surveys, and so might the age range.

FIGURE 1
Variance of EWSD respondents' raking weights across countries



daily, while the older groups report consuming only 0.82 to 1.04 grams per day, depending on the specific age bracket.

Figure 2 illustrates these differences when controlling for age and gender at the same time. Not all 16 segments in the figure are equally important; the graphs encased by the dashed line account for 80 % of consumption based on the raking results in the previous section.

TABLE 4
Regression coefficients

Variable	Coefficient (standard error)
Male	0.245*** (0.040)
Age group 18–24	0.672*** (0.065)
Age group 25–34	0.421*** (0.078)
Age group 35–44	0.448*** (0.091)
Age group 45+	0.611*** (0.116)
Frequency 4–9 days	0.233*** (0.066)
Frequency 10–19 days	0.308*** (0.063)
Frequency 20+ days	0.645*** (0.051)
Lower secondary education	−0.276*** (0.065)
Higher secondary education	−0.381*** (0.057)
College education	−0.360*** (0.059)
Income EUR 1 250–2 000	0.015 (0.050)
Income EUR 2 000–3 000	−0.095 (0.060)
Income EUR 3 000–4 000	−0.226*** (0.071)
Income EUR 4 000–5 000	−0.265*** (0.088)
Income EUR 5 000 or more	−0.186** (0.088)

N = 2 588; * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table 4 provides the coefficients and standard errors of the calculation of daily average consumption for each user in the EWSD sample. These were regressed on categorical variables coding for all variables that tend to be correlated with consumption (gender, age, frequency, education and income).

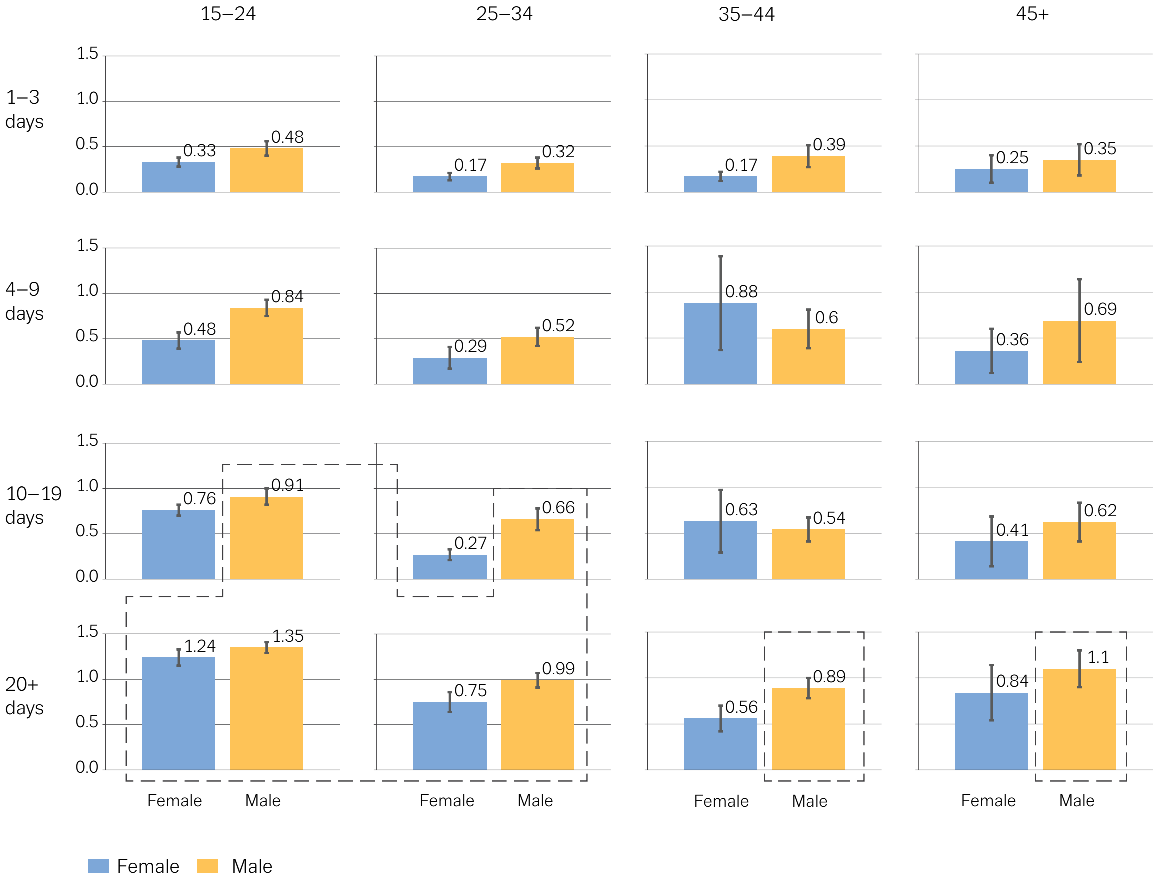
For each country, we distribute the total number of past-month users across subgroups defined by gender, age and frequency of use using GPS prevalence data. Multiplying those numbers of users by the regression model's estimates of the groups' average consumption per day of use (in grams) and the groups' past-month days of use, and also multiplying by 12 (to annualise results), gives an estimate of the country's total annual cannabis consumption.

Comparing raking and regression

We applied both methods to six countries that participated in the 2016 wave of the EWSD. Figure 3 shows that in three of the six countries, the point estimates of the two methods agree to within a difference of 5–20 %. However, the results were very different for France, where the raking method estimated the per capita consumption to be 1.8 times larger than did the regression-based approach.

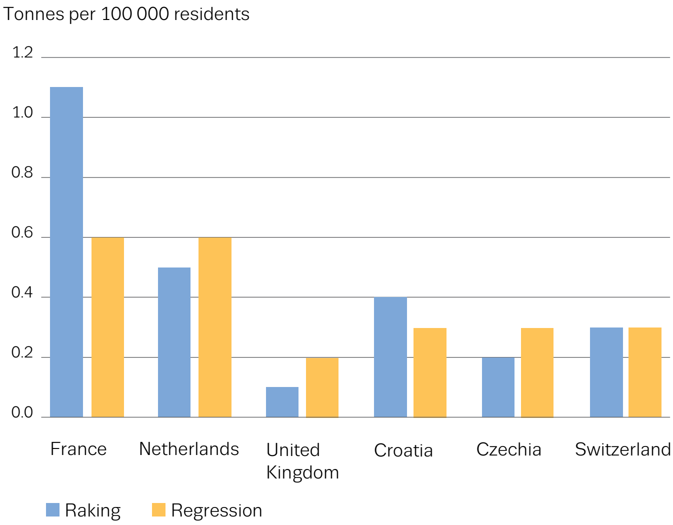
Two factors account for many of the differences. One is that raking was implemented separately in each country, whereas the regression analysis combined the data for all countries in a single expression and applied this expression to each country individually. In future analyses it would be useful to also run the regressions separately and/or include controls for country and compare the results. The second factor is that raking had

FIGURE 2
Quantity in grams per gender, age and frequency group



Error bars = 95 % CI.
The graphs encased by the dashed line account for 80 % of consumption based on the raking results in the previous section.

FIGURE 3
Point estimates of per capita consumption for two different methods



to assign high weights to some respondents in countries with large differences between the distributions of respondents in the EWSD compared to the GPS. This makes those countries' EWSD-based totals sensitive to the answers of relatively few respondents. For example, in the United Kingdom some users received a weight of 5 or more.

Discussion

Raking has advantages in theory, but can encounter challenges in practice. For example, if there are interaction groups (intersection of age, gender and frequency of use, for instance) with no observations then there is nothing to reweight. Likewise, if there are very few web survey observations in a group that has many users according to the GPS, then raking will assign a very high weight to those few respondents. One might also wonder whether the respondents the web survey attracted are really representative of their group, or whether there is something different about them that led them to participate in the survey.

The regression approach also has advantages, including a certain transparency. It starts with familiar cross-tabs of numbers of users by subgroup and estimates a consumption rate in grams per day of use for each subgroup. Anyone can look at and understand those consumption intensities and make a judgement as to whether that regression output appears plausible. Furthermore, a country that does not conduct its own web survey can use the regression-based approach (but not raking) if it believes that its users' intensities of consumption are similar to those of users in a country that did carry out a web survey. Of course, if such a country thought the regression estimated with other countries' web surveys would not provide a suitable model for its own users, then it could choose not to use the raking or the regression approach.

If web survey respondents are non-representative in ways that are not correlated with observed characteristics, then there is a source of selection that no statistical adjustment procedure can overcome. In other words, both methods could adjust for frequent users being over-represented in web survey samples, because both the GPS and web survey ask about frequency of use; however, if (1) web survey respondents differ from the general population with respect to some trait ('X'), (2) either the web survey or the GPS did not ask about X, (3) X was not completely correlated with other observable variables, and (4) X was correlated with frequency and amounts consumed per day of use, then these methods could not uncover and correct for that hidden bias. That is not a trivial limitation; many such traits can exist, but applying these methods could still be preferable to making assessments of amounts consumed per person based on simple rules of thumb or having no estimates at all.

While it is helpful that there are two ways of estimating drug market size in terms of the quantities of drugs consumed, and that those methods can agree, it would be more reassuring if these methods could be compared to a gold standard. Caulkins et al. (2019) took a modest step in that direction by comparing these methods' estimates of cannabis spending and consumption in the state of Washington, where cannabis production and sale is legal, to the figures recorded in the state's 'seed-to-sale' tracking database (which includes essentially the universe of all licensed sales, albeit not sales in the residual illicit market).

Conclusion

This study describes two approaches, namely raking and regression, for bringing together the complementary strengths of general population surveys (GPS) and web surveys to estimate the size of drug markets in terms of the quantities of drugs consumed. These methods were then applied to data collected from the six countries participating in the first wave of the European Web Survey on Drugs (EWSD) in 2016, focusing specifically on cannabis consumption.

The point estimates of per capita consumption for raking and regression agreed to within a difference of 5–20 % for all countries except France. Two factors account for many of the differences, namely that raking was implemented separately in each country while the regression analysis combined the data for all countries in a single expression and applied this expression to each country individually. While both of these methods have their own limitations, they could still represent an improvement over making projections based on simple rules of thumb or having no estimates at all.

Understanding the strengths and limitations of these methods and, hence, their potential to estimate the quantities of drugs consumed for different drug types, would benefit from occasionally validating the GPS data with biological tests (e.g., urinalysis) and improving the science on obtaining accurate information about quantities (Kilmer and Pacula, 2017). This may become especially important as consumers move beyond the use of herbal and resin forms of cannabis to include various extract-based products and edibles. That may necessitate more sophisticated ways of aggregating the consumption of different types of cannabis products.

References

Battaglia, M. P., Izrael, D., Hoaglin, D. C. and Frankel, M. R. (2009), 'Practical considerations in raking survey data', *Survey Practice* 2(5), pp. 1–10.

Caulkins, J. P., Davenport, S., Doanvo, A., Furlong, K., Siddique, A., Turner, M. and Kilmer, B. (2019), 'Triangulating web & general population surveys: Do results match legal cannabis market sales?', *International Journal of Drug Policy*, doi:10.1016/j.drugpo.2019.06.010.

Dever, J. A., Rafferty, A. and Valliant, R. (2008), 'Internet surveys: Can statistical adjustments eliminate coverage bias?', *Survey Research Methods* 2(2), pp. 47–60.

Dujourdy, L. and Besacier, F. (2017), 'A study of cannabis potency in France over a 25 years period (1992–2016)', *Forensic Science International* 272, pp. 72–80.

EMCDDA (European Monitoring Centre for Drugs and Drug Addiction) (2012), *Cannabis production and markets in Europe*, Publications Office of the European Union, Luxembourg (https://www.emcdda.europa.eu/publications/insights/cannabis-market_en).

EMCDDA (2017a), *Statistical Bulletin 2017*: prevalence of drug use, http://www.emcdda.europa.eu/data/stats2017/gps.

EMCDDA (2017b), *Statistical Bulletin 2017*: price, purity and potency, http://www.emcdda.europa.eu/data/stats2017/ppp_en.

Kilmer, B. and Pacula, R. L. (2017), 'Understanding and learning from the diversification of cannabis supply laws', *Addiction* 112(7), pp. 1128–1135.

Kilmer, B., Caulkins, J., Pacula, R. and Reuter, P. (2011), 'Bringing perspective to illicit markets: Estimating the size of the U.S. marijuana market', *Drug and Alcohol Dependence* 119, pp. 153–160.

Kilmer, B., Everingham, S. M. S., Caulkins, J., Midgette, G., Pacula, R., Reuter, P., Burns, R., Han, B. and Lundberg, R. (2014), *What America's users spend on illegal drugs, 2000–2010*, Executive Office of the President, Washington, DC.

Loosveldt, G. and Sonck, N. (2008), 'An evaluation of the weighting procedures for an online access panel survey', *Survey Research Methods* 2(2), pp. 93–105.

Rhodes, W., Dyous, C., Hunt, D., Luallen, J., Callahan, M. and Subramanian, R. (2012), *What America's users spend on illegal drugs, 2000–2006*, prepared for the Office of National Drug Control Policy, Executive Office of the President, Washington, DC.

van Laar, M., Frijns, T., Trautmann, F. and Lombi, L. (2013), 'Cannabis market: user types, availability and consumption estimates', in: Trautman, F. et al. (editors), *Further Insights into Aspects of the Illicit EU Drugs Market*, Publications Office of the European Union, Luxembourg, pp. 73–182.

Yeager, D. S., Krosnick, J. A., Chang, L., Javitz, H. S., Levendusky, M. S., Simpser, A. and Wang, R. (2011), 'Comparing the accuracy of RDD telephone surveys and internet surveys conducted with probability and non-probability samples', *Public Opinion Quarterly* 75(4), pp. 709–747.

About the EMCDDA

The European Monitoring Centre for Drugs and Drug Addiction (EMCDDA) is the central source and confirmed authority on drug-related issues in Europe. For over 25 years, it has been collecting, analysing and disseminating scientifically sound information on drugs and drug addiction and their consequences, providing its audiences with an evidence-based picture of the drug phenomenon at European level. Based in Lisbon, the EMCDDA is one of the decentralised agencies of the European Union.

About this series

EMCDDA Insights are topic-based reports that bring together current research and study findings on a particular issue in the drugs field. This paper is published as part of *Monitoring Drug Use in the Digital Age: Studies in Web Surveys*, an EMCDDA Insights that provides an overview of current knowledge and the latest developments in the field of web surveys on drug topics. The Insights contains in-depth reports on the methodology of web surveys, the available studies being carried out in different drug topics and analyses of the European Web Survey on Drugs. The Insights will be of interest to researchers and scientists, people who use drugs, policymakers and their advisors, specialists and practitioners, and all those concerned with the issue of drugs and innovative methods.

EMCDDA project group: João Matias, Alexander Soderholm, Katerina Skarupova, André Noor and Jane Mounteney.

Recommended citation: Caulkins, J. P., Furlong, K., McComas, B. Domingues, M. N., Siddique, A. and Kilmer, B. (2022), 'Comparing two approaches for estimating cannabis consumption: raking and regression', in *Monitoring drug use in the digital age: Studies in web surveys,* EMCDDA Insights (https://www.emcdda.europa.eu/publications/insights/web-surveys/comparing-approaches-estimating-cannabis-consumption-raking-regression_en).

Luxembourg: Publications Office of the European Union

PDF ISBN 978-92-9497-801-1 ISSN 2314-9264 doi:10.2810/51138 TD-XD-22-005-EN-N

This publication is only available in electronic format.

EMCDDA, Praça Europa 1, Cais do Sodré, 1249-289 Lisbon, Portugal
Tel. (351) 211 21 02 00 | info@emcdda.europa.eu
emcdda.europa.eu | twitter.com/emcdda | facebook.com/emcdda

Publications Office of the European Union