Project to explore potential approaches to missing data on problem drug use in EU27 and Norway, including interpolation and / or qualitative information based on available estimates and other routinely collected data.

Gordon Hay
June 2013
EMCDDA project responsible: Danica Thanki

Centre for Public Health, Liverpool John Moores University and
Glasgow Prevalence Estimation Limited

Correspondence should be addressed to Dr Gordon Hay at the Centre for Public Health, Liverpool John Moores University

Telephone: 0151 231 4385
Fax: 0151 231 4552
E-mail: g.hay@ljmu.ac.uk

Glasgow Prevalence Estimation

1. Introduction

This report outlines the results of a small study that sought to examine whether interpolation methods, such as regression models, can be used to provide PDU prevalence estimates for years where prevalence estimates are not readily available. The aims of the study were to

- explore routinely collected data which are potentially useful for the task in question
- describe and critically assess the methods used in interpolation
- describe any main problems if any and solutions
- make recommendations for data collection and reporting

The purpose of this exercise was to explore the possibilities to improve trends analysis, as well as to see whether it would be possible to simplify the work of national focal points by suggesting less frequent indirect methods-based estimation studies of PDU, supplemented by annual interpolation of estimates based on routinely collected data from multiple indicators.

Within this study, interpolation is considered as an approach to interpolate across time, rather than the more commonly used method known as the multivariate indicator model (MIM) which extrapolates over geographical areas, for example to construct national prevalence estimates from local prevalence estimates.

2. Data

In this section we identify routine trend data from indicators that can be used for interpolation when estimating PDU prevalence. These data include data on the other four key indicators, such as treatment demand data and drug-related deaths, and other important data sets, such as law enforcement drug seizures data and drug law offences data. The EMCDDA collates a wide range of indicator data. These data are supplied by the network of National Focal Points through a series of standard tables and much of the data are available on the EMCDDA website. In order for a table to be useful for this project, it had to include data for more than one member state and for more than one year.

There were three main sets of data considered for this project. These were:

1) A complete set of indicator data that has time series information for more than one member state and more than one year (i.e. all data in the EMCDDA Statistical Bulletin that fit those criteria).
2) Drug-related death, seizures, drug law offences and treatment demand data that were available on the EMCDDA website or could be readily derived from the available data.
3) Data derived from internal re-analyses of indicator data by the EMCDDA.

In terms of the number of countries where such analyses can be carried out this could be countries that contribute to the EMCDDA datasets or any subset of them. We specifically examine the data for a set of countries for which the EMCDDA already published problem opiate use time trend tables as of the 2012 Annual report on the state of the drugs problem in Europe:

- Austria
- Cyprus
- Czech Republic
- Germany
- Spain
- Greece
- Italy
- Malta
- Slovakia

Complete set of indicator data

Box 1 lists the tables from the EMCDDA website that were considered as being potentially suitable for interpolation.

BOX 1 Tables from EMCDDA website considered for use in interpolation

Table DLO-1
Drug law offences, 1995 to 2010; Part (i) Number of reports of offences

Table DLO-1
Drug law offences, 1995 to 2010; Part (ii) Number of reports of persons

Table DLO-4
Drug law offences related to drug use or possession for use; 2003 to 2010, Part (i) Number and percentage

Table DLO-7
Heroin-related offences, 2003 to 2010; Part (i) Number and percentage of all drug law offences

Table DLO-8
Cocaine-related offences, 2003 to 2010; Part (i) Number and percentage of all drug law offences

Table DLO-109
Number of reports for drug law offences, 1985 to 2010; Part (i) Number of reports of offences

Table DLO-109
Number of reports for drug law offences, 1985 to 2010; Part (ii) Number of reports of persons

Table DRD-2
Number of drug-induced deaths recorded in EU Member States according to national definitions; Part (i) Total drug-induced deaths, 1995 to 2010

Table DRD-3
Number of drug-induced deaths recorded in EU Member States and Norway according to EMCDDA standard definition 'Selection B', 1995 to 2010

Table DRD-4
Number of drug-induced deaths recorded in EU Member States according to EMCDDA standard definition 'Selection D', 1995 to 2010

Table DRD-107
Number of drug-induced deaths recorded in EU Member States according to national definitions; Part (i) Total drug-induced deaths, 1985 to 2010

Table SZR-7
Number of heroin seizures 1995 to 2010; Part (i) 1995 to 2010

Table SZR-9
Number of cocaine seizures 1995 to 2010

Table SZR-10
Quantities (kg) of cocaine seized 1995 to 2010

Table TDI-2
Clients entering treatment and reporting treatment units, 1998 to 2010; Part (ii) All clients by country and year of treatment

Table TDI-2
Clients entering treatment and reporting treatment units, 1998 to 2010; Part (ii) All clients by country and year of treatment

TDI, SZR, DLO and DRD data from website

From the wider group of datasets that were potentially useful for interpolating within countries over time, it was decided that treatment demand data (TDI), seizures (SZR), drug law offences (DLO) and drug-related death data (DRD) were the most appropriate to focus on. The TDI data were constructed by combining the TDI outpatient data with the percentage that were using opiates. For those analyses the data were for the years 2005 to 2010 (6 consecutive years). The drug-related deaths data in this section relate to all drug-related deaths, not those that were specifically related to opiate use.

Internal re-analyses of EMCDDA data

Two datasets were specifically requested from the EMCDDA as part of this project. These were a set of drug-related death data which specifically related to opiate use and a treatment demand dataset that provided information on the number of reported treatments for heroin use and the number of reported treatments for opiate use. These were augmented by drug law offence data and the seizure data that is described above.

3 Methods

Within this study, interpolation is taken to mean fitting a linear regression model where the problem drug use (or problem opiate use) prevalence rate is the dependent variable and the available indicator data are the independent variables. This is commonly known as the multivariate indicator model (or multiple indicator model) when extrapolating across geographical areas, e.g. to get a national estimate when only a limited number of local-level estimates are available for a country. When there is only one indicator the multiple indicator method is similar to a multiplier method (e.g. the mortality multiplier or a treatment multiplier) and will be exactly the same as a multiplier method if the regression model is forced to have a zero intercept (i.e. in the case of a treatment multiplier if the number in treatment is zero then the prevalence must be zero). Once a regression model has been established, new estimates can be interpolated by either entering the indicator data for the new time period into the model, or using the relevant procedures in a statistical package which, in addition to produce estimates, can also be used to derive relevant confidence intervals for the estimates.

In this study we are looking at interpolating across time, i.e. if there are prevalence estimates available for 2 or more years is it possible to fit a regression model and predict prevalence for a year when an actual estimate was not available.

There is a substantial amount of information about regression models in the scientific literature, including methods of identifying whether the regression model adequately fits the existing data. Two related statistical measures can be used in deciding whether the regression model provides a good fit, the R^2 and the adjusted R^2. The adjusted R^2 accounts for the issue that increasing the number of independent variables in a regression model can only increase the goodness of fit, and therefore favours simpler, more parsimonious, models.

A more practical issue for any regression model when using indicator data to predict prevalence is that the regression model should have a slope that seems appropriate to the indicator, for example if the indicator value increases then you would expect prevalence to also increase. As an example, it would be expected that the prevalence of problem drug use would increase if the number of drug-related deaths increases. Similarly if the number of drug users in treatment increases then prevalence should also increase. While there can be reasons why this would not be the case, i.e. increasing treatment coverage, if that was the case then perhaps regression models should not be used to interpolate or alternative regression models, such as those that would take into account a time lag between commencement of drug use and entering treatment, should be explored.

In all of the analyses described in this report, the regression models regress problem opiate use (expressed as a rate) and indicator data, also expressed as a rate.

4 Results

TDI, SZR, DLO and DRD data from website

Since there were more data on problem opiate use (POU) a decision was taken to consider it as the 'prevalence' estimates as those data may have the best chance of correlating against indicator data. The data that referred to cocaine and amphetamines use were therefore not considered further.

The data were regressed against the variables that are available as a time series and that were more likely to be the most correlated, the drug-related death indicator data, seizures data, drug law offence data and the treatment demand indicator. In the first instance a treatment demand opiate use data series was derived by combining the TDI data with the percentage that are using opiates (in the outpatients data). In the first instance the analyses only considered the nine countries that had the most complete series of problem opiate use estimates.

To see if there is any possibility of using interpolation across countries the following two scatterplots were created.

Figure 1 is a scatterplot between the problem opiate use estimates against the drug-related deaths indicator data.

Figure 1 Scatterplot of problem opiate use data against drug-related death data for nine countries

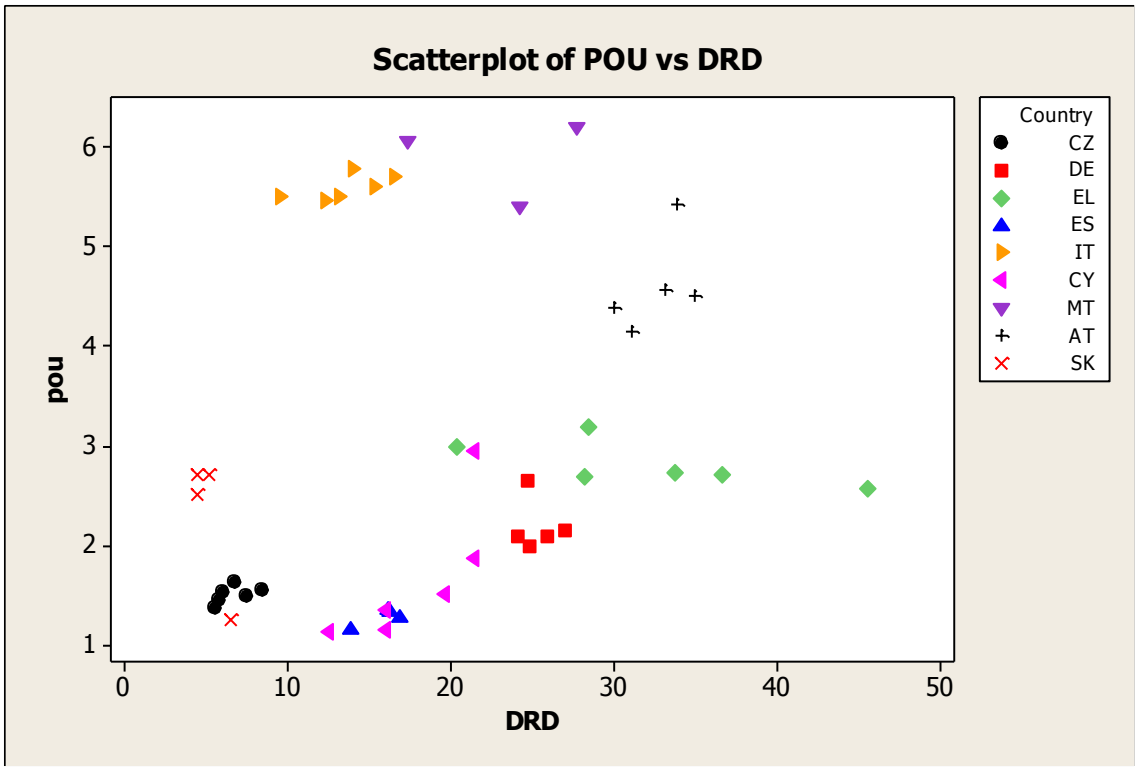
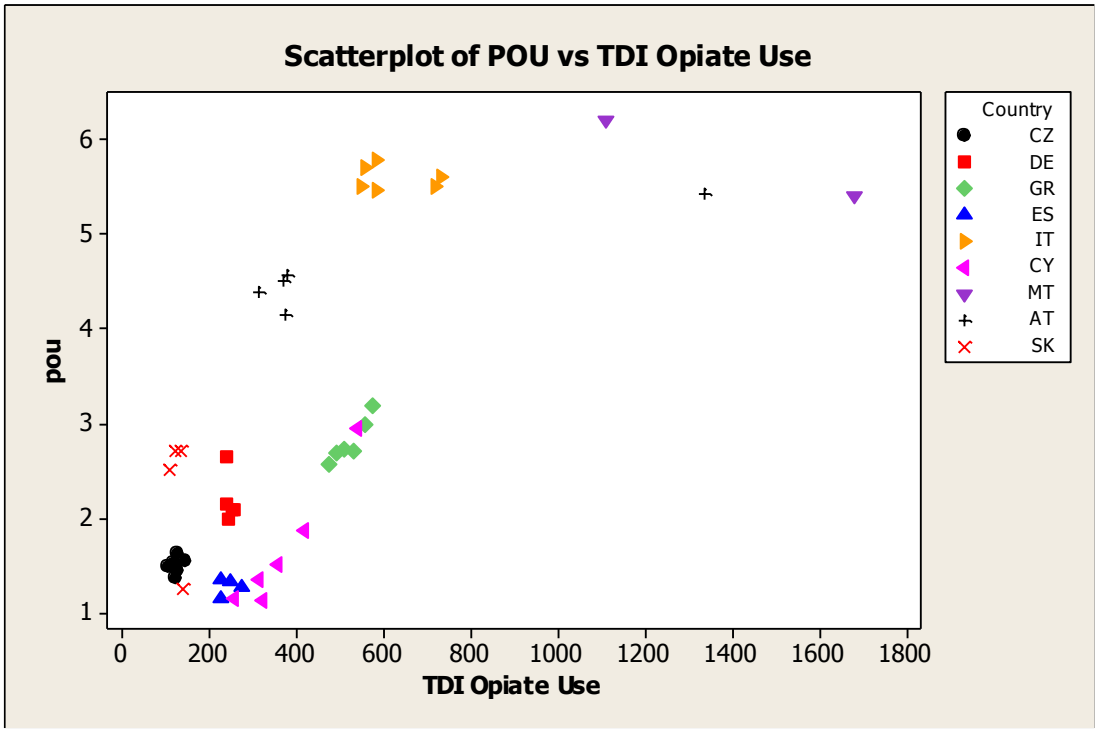


Figure 1 suggests that there is no correlation across countries, i.e. the drug-related death rate in one country is not useful in predicting problem opiate use in another country. This is the same for POU against TDI as seen in Figure 2.

Figure 2 **Scatterplot of problem opiate use data against treatment demand indicator data for nine countries**



Note: There is an outlier (not shown in Figure 2) for Malta at approximately 4000 on the X axis (TDI Opiate Use) and 6 on the Y axis (POU)

We can look at correlations within countries and there are mixed results. The regression results (including R^2 and adjusted R^2 values) as listed in the following tables, first for POU against DRD, then for POU against TDI, then regression models which regress POU against DLO or SZR.

Table 1 **Summary of regression analyses POU v DRD (2005 – 2010)**

Country	Cases	R^2	Adjusted R^2	Regression Model
Austria	5	27.5	3.4	POU = 0.51 + 0.126 x DRD
Cyprus	6	58.2	47.7	POU = - 0.97 + 0.147 x DRD
Czech Republic	6	28.5	10.6	POU = 1.21 + 0.0433 x DRD
Germany	5	3.1	0.0	POU = 3.22 - 0.041 x DRD
Greece	6	47.6	34.5	POU = 3.41 - 0.0189 x DRD
Italy	6	37.7	22.1	POU = 5.17 + 0.0312 x DRD
Malta	3	0.1	0.0	POU = 5.96 - 0.0031 x DRD
Slovakia	4	79.8	69.7	POU = 5.55 - 0.648 x DRD
Spain	4	61.7	42.5	POU = 0.453 + 0.0521 x DRD

Table 2 Summary of regression analyses POU v TDI (2005 – 2010)

Country	Cases	R^2	Adjusted R^2	Regression Model
Austria	5	89.0	85.3	POU = 4.01 +0.00106 x TDI
Cyprus	6	93.9	92.3	POU = - 0.805 + 0.00664 x TDI
Czech Republic	6	4.5	0.0	POU = 1.30 + 0.00160 x TDI
Germany	5	14.5	0.0	POU = 5.28 - 0.0124 x TDI
Greece	6	87.5	84.3	POU = - 0.177 + 0.00570 x TDI
Italy	6	4.9	0.0	POU = 5.80 - 0.000338 x TDI
Malta	3	3.0	0.0	POU = 5.78 + 0.000051 x TDI
Slovakia	4	22.6	0.0	POU = 5.45 - 0.0245 x TDI
Spain	4	1.7	0.0	POU = 1.15 + 0.00050 x TDI

There does not seem to be enough correlation for either DRD or TDI for Czech Republic, Germany or Malta (there was not enough data for Malta to run the regression on DRD and TDI) to allow for any appropriate interpolation. There does not seem to be sufficient correlation between POU and DRD / TDI in Italy, but the results for Greece, Spain, Cyprus, Austria and Slovakia are much more positive.

In Table 3 we can look at the correlation between the problem opiate use estimates and data on the number of heroin seizures (from standard table SZR-7), expressed as a rate per 10,000 population.

Table 3 Summary of regression analyses POU v SZR (2005 – 2010)

Country	Cases	R^2	Adjusted R^2	Regression Model
Austria	5	90.9	87.9	POU = 3.03 + 0.958 x SZR
Cyprus	5	0.1	0.0	POU = 1.75 - 0.047 x SZR
Czech Republic	6	1.9	0.0	POU = 1.56 – 0.51 x SZR
Germany	5	5.6	0.0	POU = 0.09 + 1.73 x SZR
Greece	6	2.8	0.0	POU = 3.01 - 0.036 x SZR
Italy	6	56.7	45.9	POU = 7.68 – 2.27 x SZR
Malta	3	2.8	0.0	POU = 5.50 + 0.22 x SZR
Slovakia	4	33.9	0.9	POU = 8.62 - 10.9 x SZR
Spain	4	98.7	98.0	POU = 1.78 – 0.196 x SZR

Again the results are mixed, with good correlation in Austria and Spain and moderate correlation in Italy. In Spain and Italy the direction of the slope in the regression model suggests that seizures decrease when prevalence increases, which may be counter-intuitive. In Table 4 we look at the results from the regression analyses that regress problem opiate use against the heroin drug law offence data.

Table 4 Summary of regression analyses, POU against Heroin DLO

Country	Cases	R^2 (%)	Adjusted R^2 (%)	Equation
Austria	5	79.7	73.0	POU = 2.55 + 0.341 x H_DLO
Cyprus	6	0.7	0.0	POU = 1.85 - 0.133 x H_DLO
Czech Republic	6	15.1	0.0	POU = 1.27 + 1.29 x H_DLO
Germany	5	73.5	64.6	POU = -1.36 + 0.659 x H_DLO
Greece	6	17.0	0.0	POU = 3.30 - 0.0622 x H_DLO
Italy	5	70.0	62.5	POU = 6.37 - 0.328 x H_DLO
Malta	3	31.4	0.0	POU = 6.92 - 0.178 x H_DLO
Slovakia	3	33.8	0.0	POU = 1.43 + 1.53 x H_DLO
Spain	4	96.7	95.1	POU = 1.86 - 0.207 x H_DLO

To examine these issues further two specific datasets were supplied by the EMCDDA, a specially constructed opiate drug-related death dataset and an opiate treatment demand dataset.

Internal re-analyses of EMCDDA data

When fitting regression models to the data specifically supplied by the EMCDDA for this project we get the following results

Table 5 Summary of regression analyses, POU against Opiate DRD (2005-2010)

Country	Cases	R^2 (%)	Adjusted R^2 (%)	Equation
Austria	5	31.4	8.5	POU = 0.55 + 0.127 x DRD
Cyprus	6	58.9	48.7	POU = -0.94 + 0.166 x DRD
Czech Republic	5	34.1	12.1	POU = 1.27 + 0.0929 x DRD
Greece	5	63.8	51.8	POU = 3.13 - 0.0173 x DRD
Italy	6	20.4	0.5	POU = 5.30 + 0.0270 x DRD
Malta	3	3.0	0.0	POU = 5.66 + 0.0115 x DRD
Slovakia	4	67.7	51.6	POU = 4.02 - 0.489 x DRD

Table 6 Summary of regression analyses, POU against Heroin TDI (2005-2010)

Country	Cases	R^2 (%)	Adjusted R^2 (%)	Equation
Austria	4	0.5	0.0	POU = 4.58 - 0.0050 x H_TDI
Cyprus	5	81.9	75.8	POU = -3.45 + 0.00871 x H_TDI
Czech Republic	5	0.2	0.0	POU = 1.41 + 0.00045 x H_TDI
Germany	5	53.6	38.1	POU = 1.41 + 0.00182 x H_TDI
Greece	5	82.7	76.9	POU = -0.163 + 0.00534 x H_TDI
Italy	5	8.5	0.0	POU = 6.09 - 0.00066 x H_TDI
Slovakia	4	10.2	0.0	POU = -0.09 + 0.0121 x H_TDI
Spain	3	81.8	63.7	POU = -0.27 + 0.00265 x H_TDI

Table 7 Summary of regression analyses, POU against Opiate TDI

Country	Cases	R^2 (%)	Adjusted R^2 (%)	Equation
Austria	4	24.7	0.0	POU = 7.42 - 0.00589 x O_TDI
Cyprus	5	10.1	0.0	POU = -10.8 + 0.00328 x O_TDI
Czech Republic	5	21.4	0.0	POU = 2.86 - 0.00380 x O_TDI
Germany	5	71.6	62.2	POU = 5.03 - 0.0500 x O_TDI
Greece	5	36.9	15.9	POU = 39.8 - 0.138 x O_TDI
Italy	5	0.9	0.0	POU = 4.52 + 0.021 x O_TDI
Slovakia	4	11.2	0.0	POU = 6.87 - 0.008 x O_TDI
Spain	3	98.8	97.6	POU = -187 + 0.0472 x O_TDI

Again the results are mixed. For Cyprus, Greece and Slovakia there may be enough correlation between the problem opiate use estimates and the opiate drug related death data to allow for interpolation however for Greece the regression model suggests that prevalence decreases when drug-related deaths increase which is perhaps counter-intuitive. For Cyprus and Greece there appears to be sufficient correlation between the problem opiate use estimates and the heroin treatment demand data. For Germany there does appear to be sufficient correlation when looking at the opiate treatment demand data; the correlation for Germany is less when looking at heroin data. There is a very high correlation between the opiate treatment demand data and problem drug use estimates for Spain, but that is likely to be an artefact of the small number of data points (3). In general, the heroin treatment demand data is more correlated with problem opiate use prevalence that the opiate treatment demand data.

Multiple regression models

The above analyses reported in Table 1 to Table 7 only fit regression models that compare problem opiate use against one indicator at a time. These analyses can be extended to ones that include more than one indicator and the only restrictions to the number of indicators that can be included would be the number of data points (or cases) where there are problem opiate use estimates and the relevant indicator data for that year. In the following analyses we can regress problem opiate use (POU) against the opiate drug-related death data (specifically obtained from the EMCDDA), the treatment demand data (TDI) (specifically obtained from the EMCDDA), the drug law offences data (DLO) and the seizures data (SZR). For the TDI data either the heroin data could be used or the opiate data, and the choice of which TDI data was used in the multiple regression models was made by selecting the regression analyses (from Table 6 or Table 7) that had the highest R^2 value for that country.

With four different indicators there are 11 different multiple regression models that could be considered. One of the models has all four indicators and there will be a set of four models that miss out one of the indicators and another six models that have two indicators in the model. With a maximum of six POU estimates in the time period 2005 to 2010, the maximum number of indicators that could be employed in a model that is not saturated (i.e. provides a perfect fit to the available data only because all of the available data is used within the model) is five. Not all countries will have a complete set of POU estimates for that time period, and not all countries with have complete sets of indicator data for that period. In particular, two of the countries in the analyses described above (Germany and Spain) did not have the relevant drug-related death data. Out of the maximum number of 99 different regression models that could be fitted (11 for each of the 9 countries) only 51 analyses had sufficient data.

Appendix 1 contains a table that summarises the results from these 51 regression analyses. Although the results are mixed, there are some analyses that have relatively high adjusted R^2 values (>90) that suggest quite a high correlation between the problem opiate use estimates and the combined indicator data. The results for Greece are particularly interesting as there are cluster of analyses that have high R^2 values, similarly for Italy. Although it was only possible to fit one multiple regression model in Spain (including DLO and SZR), that model provided a very good fit to the available data. In Slovakia it was only possible to fit three different multiple regression models and one (DRD and TDI) provided a good fit.

Supplementary analyses – England (UK)

In the analyses above, there may be issues with differing case definitions and differences in study design etc. across years. In order to look at interpolation in a situation where it would be expected that there was less variation in external factors over consecutive years a re-analyses of data from a series of problem opiate use estimation studies in England (UK) was carried out. Fuller details of these studies are available at Hay *et al* 2006 and Hay *et al* 2007.

The prevalence studies in England use the capture-recapture method (using four data sources) to estimate the prevalence of opiate use at the local level, and where there are issues using that method to estimate prevalence the multivariate indicator method is used. The study has been carried out using exactly the same study design and contributing data sources for a series of 6 years, starting with the financial year 2004/05 (1st April 2004 to 31st March 2005), with yearly estimates also for 2005/06, 2006/07, 2008/09, 2009/10 and 2010/11. The local level analyses were carried out in all 149 local areas of England, but across the 6 time periods there were 28 areas that had a problem opiate use estimate derived using the capture-recapture method for each year. Within these additional analyses the prevalence of opiate use is regressed against the data on heroin use in treatment. Table 8 summarises the regression models for each area.

Table 8 Summary of regression analyses, prevalence of opiate use against opiate treatment data for 28 areas of England

Regression model	R^2 (%)	Adjusted R^2 (%)
PREV = 96.2 - 0.650 x TREAT	18.0	0.0
PREV = 41.2 + 0.609 x TREAT	8.8	0.0
PREV = 10.2 + 1.42 x TREAT	19.3	0.0
PREV = 8.45 + 1.49 x TREAT	70.7	63.4
PREV = 129 - 0.760 x TREAT	71.7	64.6
PREV = 45.6 + 0.403 x TREAT	7.2	0.0
PREV = 56.8 + 0.18 x TREAT	0.7	0.0
PREV = 96.4 + 0.106 x TREAT	0.9	0.0
PREV = 79.1 - 0.165 x TREAT	11.2	0.0
PREV = 94.5 - 0.844 x TREAT	46.3	32.9
PREV = 41.3 + 0.118 x TREAT	0.5	0.0
PREV = 49.1 + 0.015 x TREAT	0.0	0.0
PREV = 132 - 0.32 x TREAT	2.2	0.0
PREV = 33.2 + 0.878 x TREAT	18.2	0.0
PREV = 64.1 - 0.341 x TREAT	11.6	0.0
PREV = 39.6 + 0.506 x TREAT	16.8	0.0
PREV = 23.2 + 1.08 x TREAT	29.7	12.1
PREV = 62.3 - 0.286 x TREAT	22.6	3.3
PREV = 63.5 - 0.266 x TREAT	19.9	0.0
PREV = 106 - 0.292 x TREAT	6.2	0.0
PREV = 20.4 + 1.28 x TREAT	42.1	27.6
PREV = 39.1 + 0.583 x TREAT	15.4	0.0
PREV = 85.1 - 0.89 x TREAT	10.5	0.0
PREV = 57.2 + 0.710 x TREAT	12.8	0.0
PREV = 16.6 + 0.977 x TREAT	23.9	4.8
PREV = 116 - 0.176 x TREAT	1.6	0.0
PREV = 22.2 + 0.964 x TREAT	45.7	32.1
PREV = 43.8 + 0.808 x TREAT	17.8	0.0

Out of the 28 areas, there were only two where the R^2 (and adjusted R^2) values are greater than 50%. In many areas the adjusted R^2 values are zero. In ten of the areas, the regression model has a negative slope, suggesting that prevalence decreases when the numbers in treatment increases.

We can also use data from England to explore any differences between the correlations that are found across areas (as with in the multivariate indicator model) against correlations across time (the focus of this report). We did this by randomly selecting six sets of areas out of the 28 areas that featured in Table 8 and fitting a regression model that, in addition to the treatment data, has presence or absence from each of the six areas as a variable in the model along with the treatment data. Although there may be more statistically rigorous approaches to including geographical data within a regression model, the approach of including presence or absence in an area as a binary variable in the regression model provides an indication as to whether any correlation between prevalence and treatment data is stronger across time or across area.

We did this ten times with different random selections of six areas. Table 9 presents some information from those analyses.

Table 9 **Summary of regression analyses, prevalence of opiate use against opiate treatment data and area for randomly selected sets of areas of England**

Adjusted R^2 (%)	P value (Treatment)	Other significant P values	Slope (Treatment)
82.4	0.35	0.46	-0.269
78.4	0.24	None	-0.372
82.2	0.56	0.88	-0.131
75.9	0.15	0.78,0.56,0.47	0.41
38.8	0.22	0.36,0.14	0.419
90.4	0.28	0.49	-0.242
51.6	0.58	0.62	0.168
88.7	0.40	none	0.169
64.4	0.16	none	0.398
89.5	0.37	none	-0.204

The lowest adjusted R^2 value was 38.8%, however most were over 75%. In none of the ten analyses was the p value for the treatment coefficient in the regression model significant (less than 0.05). In other words, in each of these ten simulated analyses the numbers in treatment were not seen as statistically significant predictors of prevalence. When looking at the variables that represented the area, in four of the ten simulated analyses all of these geographical variables were statistically significant, in a further four only one geographical variable was not significant and in the remaining analysis three out of the six variables representing area were not significant. In half of these simulated analyses the slope suggests that prevalence decreases as the numbers in treatment increases, which may be counter-intuitive. These simulated analyses suggest that when considering variation across time and geographical area then changes across geographical area are far more correlated with prevalence than across time.

5 Discussion and recommendations

The analyses described in this report have considered whether the prevalence of problem drug use (specifically problem opiate use) is sufficiently correlated with readily available indicator data such as data on treatment (particularly treatment demand indicator data), drug-related death data or seizures data.

Although there were a few instances where it does appear that there is sufficient correlation between the problem opiate estimates and available indicator data, these instances were in the minority when only considering one indicator at a time. More promising results were found when using multiple regression methods but only a few countries had sufficient data (given the data sources explored) to use such a method. In addition, when looking at analyses for England (where in the multivariate indicator method analyses across geographical areas there was consistently quite a high correlation between prevalence and treatment data) there were few areas where there appeared to be sufficient correlation between prevalence and treatment data over time. Further analyses of the English data suggested that when fitting a regression model where prevalence of opiate use is the dependent variable and treatment data is the independent variable (with area included as a series of presence / absence binary variables) then the treatment data was not seen as statistically significant in the regression model.

There are several caveats that need to be considered. The first main caveat is that although the available indicator data is apparently related to problem drug use prevalence, the real relationships may be far more complex. In the case of treatment data, although it may be expected that numbers in treatment would be correlated with underlying prevalence (and

many prevalence estimation methods, such as the treatment multiplier or a multivariate multiple indicator model including treatment, are explicitly based on this relationship) there may be other issues. In particular, there is usually a time gap between someone beginning to use drugs (or coming under the definition of problem drug use) and entering treatment. Many of the available methods for estimating the incidence of problem drug use are based on this very time-lag. Thus if prevalence varies between higher and lower levels, the treatment data may not be reflecting changes in prevalence at that time, rather it would (in large part) reflect changes in historical prevalence. This would be quite a serious issue when using interpolation to estimate changes in prevalence over time, yet would not be so much of a problem when using the multivariate indicator method to extrapolate across areas as it may be appropriate to assume that any temporal lag may be more similar across geographical areas within the same country, notwithstanding any differences in treatment uptake.

For the EMCDDA TDI data, there is also the additional issue that the data series reflects the demand for treatment (i.e. new treatment episodes) rather than the 'treatment prevalence' (although these data are being collected for some countries and may therefore be of use in any future extrapolation analyses). If there were high levels of the drug using population already in treatment then treatment demand or 'incidence' could be low. Treatment incidence could also be unduly influenced by treatment coverage such that if the availability of treatment then the numbers entering treatment could be increasing, even if the prevalence of problem drug use was not increasing.

The supplementary analyses that were carried out using data from England should be relatively robust to this as treatment data used in the analyses was derived from a national register that includes all individuals in structured treatment (not just new treatment demands) and it can possibly be assumed that treatment coverage was relatively stable over the period under study. Yet still in these analyses the correlation between opiate use prevalence and opiate use treatment data across yearly time periods was relatively weak suggesting that it would be wrong to estimate or predict prevalence using interpolation.

There may also be other issues with prevalence estimates and indicator data. In the case of the indicator data there may be fluctuations that, although possibly due to underlying issues, appear to be almost random. For example, it may seem credible that problem drug use prevalence may be highly correlated with drug-related death data (if case definitions etc. are consistent) yet most drug-related death time series data do seem to fluctuate, with these fluctuations not actually reflecting changes in underlying prevalence. It has to be mentioned also that other factors, with potentially very strong influence, can play a role in the drug-related death rate, mainly the variation in drug purities at the street level.

Also, police seizures do not reflect only real trends in the prevalence of drug consumption but also changing practices of the police and sometimes changes in drug laws.

Moreover, some even simpler, core issues related to all the data sets used, include changes in the reporting systems over time, which may affect the reflection of real trends in the data and imperfect coverage – the coverage of reporting varies from country to country and over time and there are a lot of cases where it is not 100%. Many of the multiple regression models discussed in this report are affected by missing data. There may be merit in exploring these models further once the coverage of reporting improves or when more problem opiate use estimates become available.

All these factors have to be examined case by case at the country level in order to have a better insight into the validity or theoretical appropriateness of this type of interpolation. Moreover, this report has only dealt with data usually available at the EU level. There might be other useful data sources available at the country level.

In the case of the prevalence estimates, it does have to be re-iterated that the best available information on prevalence does come from estimates. Aside from the problem that prevalence estimates are subject to issues about case definition and differences in methods used, the estimates should be considered alongside the confidence intervals and any changes in prevalence across yearly time periods could be more to do with the error within the estimation process (which would have been accounted for by the confidence interval) rather than true changes in prevalence. Thus when the best available prevalence estimates are regressed

against the best available indicator data (particularly at yearly intervals) it may be that the regression analyses are only really comparing one apparently random looking series against another randomly looking series. These issues may be resolved by trying to incorporate confidence intervals into any regression model used to interpolate prevalence, or use more advanced statistical approaches to analysing time series data.

A further caveat is that a general lack of correlation between the indicator data and the available prevalence estimates does not definitely mean that the indicator data should not be used to predict prevalence. If, for example, the lack of correlation between prevalence and indicator data is primarily due to issues about not including or accounting for the statistical uncertainty of the actual estimates then it may be that an estimate derived by interpolation using indicator data would offer estimates that are less affected by statistical uncertainty (although those resultant estimates would be subject to their own statistical uncertainty that would be hard to measure). In other words, extrapolated estimates derived using regression models and available indicator data may give a smoother, more realistic time trend rather than a series of individual, disparate prevalence estimates produced using statistical methods such as the mortality multiplier method or capture-recapture.

In summary, there appears to be little support for recommending that, in general, interpolation should be relied on to provide estimates for years where specific prevalence estimation studies have not been carried out. The more positive results in some countries does, however, suggest that for some countries there may be merit in looking in more detail at the strengths and weaknesses of the available indicator data and the relationship between those data and the available prevalence estimates.

6 References

Hay, G, Gannon, M, MacDougall, J, Millar, T, Eastwood, C and McKeganey, N. (2006) Local and national estimates of the prevalence of opiate use and / or crack cocaine use (2004/05) in Singleton, N, Murray, R and Tinsley, L. Measuring different aspects of problem drug use: methodological developments. Home Office Online Report 16/06, Available: http://www.homeoffice.gov.uk/rds/pdfs06/rdsolr1606.pdf [08/04/2008].

Hay, G, Gannon, M, MacDougall, J, Millar, T, Eastwood, C and McKeganey, N. (2007) National and regional estimates of the prevalence of opiate use and/or crack cocaine use 2005/06: a summary of key findings. Home Office Online Report 21/07, Available: http://www.homeoffice.gov.uk/rds/pdfs07/rdsolr2107.pdf [08/04/2008].

Appendix 1

Table A1 Summary of multiple regression analyses (2005 to 2010)

Country	Cases	R^2	Adjusted R^2	Indicators	Regression Model
Austria	5	94.4	77.5	DLO, SZR, DRD	POU = 4.69 - 0.28 x DLO + 1.79 x SZR - 0.04 x DRD
Austria	5	92.6	85.3	DLO, SZR	POU = 3.41 - 0.20 x DLO + 1.48 x SZR
Austria	5	80.1	60.3	DLO, DRD	POU = 2.09 + 0.32 x DLO + 0.02 x DRD
Austria	4	55.2	0.0	DLO, TDI	POU = 12.8 - 0.33 x DLO - 0.01 x TDI
Austria	5	91.6	83.2	SZR, DRD	POU = 3.69 + 1.03 x SZR - 0.02 x DRD
Austria	4	71.8	15.4	SZR, TDI	POU = - 27.7 + 4.65 x SZR + 0.05 x TDI
Austria	4	51.3	0.0	DRD, TDI	POU = - 2.1 + 0.10 x DRD + 0.007 x TDI
Cyprus	5	88.3	53.2	DLO, SZR, DRD	POU = - 1.26 - 4.40 x DLO + 4.33 x SZR + 0.20 x DRD
Cyprus	5	95.2	80.8	DLO, DRD, TDI	POU = - 4.80 + 0.36 x DLO + 0.067 x DRD + 0.008 x TDI
Cyprus	5	12.6	0.0	DLO, SZR	POU = 1.84 - 4.77 x DLO + 4.67 x SZR
Cyprus	5	60.0	33.3	DLO, DRD	POU = - 0.71 - 0.17 x DLO + 0.17 x DRD
Cyprus	5	88.4	76.7	DLO, TDI	POU = - 5.40 + 0.55 x DLO + 0.01 x TDI
Cyprus	5	77.6	55.2	SZR, DRD	POU = - 1.38 - 0.03 x SZR + 0.21 x DRD
Cyprus	4	87.7	63.1	SZR, TDI	POU = - 5.47 + 0.59 x SZR + 0.01 x TDI
Cyprus	5	92.7	85.4	DRD, TDI	POU = - 3.54 + 0.08 x DRD + 0.007 x TDI
Czech Republic	5	93.6	74.6	DLO, SZR, DRD	POU = 1.07 + 4.51 x DLO - 4.61 x SZR + 0.08 x DRD
Czech Republic	5	91.1	64.4	DLO, SZR, TDI	POU = 2.85 + 4.12 x DLO - 5.17 x SZR - 0.004 x TDI
Czech Republic	6	74.4	57.4	DLO, SZR	POU = 1.23 + 4.82 x DLO - 4.82 x SZR
Czech Republic	5	41.1	0.0	DLO, DRD	POU = 1.13 + 0.88 x DLO + 0.09 x DRD
Czech Republic	5	30.3	0.0	DLO, TDI	POU = 1.59 + 2.31 x DLO - 0.001 x TDI
Czech Republic	5	37.3	0.0	SZR, DRD	POU = 1.35 - 0.64 x SZR + 0.09 x DRD
Czech Republic	5	65.0	30.0	SZR, TDI	POU = 4.69 - 4.20 x SZR - 0.007 x TDI
Czech Republic	4	51.3	0.0	DRD, TDI	POU = - 1.97 + 0.23 x DRD + 0.008 x TDI
Germany	5	78.7	15.0	DLO, SZR, TDI	POU = 1.1 + 0.70 x DLO - 1.95 x SZR - 0.005 x TDI
Germany	5	78.7	57.4	DLO, SZR	POU = 0.44 + 0.76 x DLO - 1.96 x SZR
Germany	5	73.6	47.2	DLO, TDI	POU = - 0.0 + 0.53 x DLO - 0.01 x TDI
Germany	5	75.4	50.9	SZR, TDI	POU = 7.37 - 1.64 x SZR - 0.06 x TDI
Greece	5	99.4	97.5	DLO, SZR, DRD	POU = 5.03 - 0.88 x DLO + 0.87 x SZR + 0.0005 x DRD
Greece	5	99.4	97.5	DLO, SZR, TDI	POU = 5.27 - 0.91 x DLO + 0.88 x SZR - 0.0003 x TDI
Greece	5	95.6	82.3	DLO, DRD, TDI	POU = - 21.7 + 0.77 x DLO + 0.08 x DRD + 0.03 x TDI
Greece	5	97.8	91.3	SZR, DRD, TDI	POU = - 9.64 + 0.42 x SZR + 0.05 x DRD + 0.02 x TDI

Country	Cases	R^2	Adjusted R^2	Indicators	Regression Model
Greece	6	58.4	30.6	DLO, SZR	POU = 3.06 - 0.33 x DLO + 0.41 x SZR
Greece	4	80.0	60.0	DLO, DRD	POU = 4.18 - 0.12 x DLO - 0.02 x DRD
Greece	5	85.4	70.7	DLO, TDI	POU = - 0.80 + 0.05 x DLO + 0.006 x TDI
Greece	5	74.2	48.4	SZR, DRD	POU = 3.86 - 0.11 x SZR - 0.02 x DRD
Greece	5	87.0	74.0	SZR, TDI	POU = - 0.56 + 0.06 x SZR + 0.005 x TDI
Greece	5	84.1	68.2	DRD, TDI	POU = 0.44 - 0.004 x DRD + 0.004 x TDI
Italy	6	89.3	73.2	DLO, SZR, DRD	POU = 7.57 - 0.25 x DLO - 1.50 x SZR - 0.002 x DRD
Italy	5	100.0	99.9	DLO, SZR, TDI	POU = 7.80 - 0.29 x DLO - 1.00 x SZR - 0.0008 x TDI
Italy	5	95.9	83.6	DLO, DRD, TDI	POU = 7.00 - 0.32 x DLO + 0.023 x DRD - 0.001 x TDI
Italy	5	55.8	0.0	SZR, DRD, TDI	POU = 7.74 - 2.11 x SZR - 0.001 x DRD - 0.0002 x TDI
Italy	6	89.2	82.0	DLO, SZR	POU = 7.52 - 0.24 x DLO - 1.46 x SZR
Italy	6	72.6	54.4	DLO, DRD	POU = 6.20 - 0.30 x DLO + 0.01 x DRD
Italy	5	91.6	83.2	DLO, TDI	POU = 7.21 - 0.35 x DLO - 0.001 x TDI
Italy	6	57.5	29.1	SZR, DRD	POU = 7.48 - 2.13 x SZR + 0.006 x DRD
Italy	5	55.8	11.6	SZR, TDI	POU = 7.71 - 2.08 x SZR - 0.0003 x TDI
Italy	5	32.5	0.0	DRD, TDI	POU = 5.80 + 0.05 x DRD - 0.001 x TDI
Slovakia	4	76.7	30.1	SZR, DRD	POU = - 1.68 + 11.8 x SZR - 0.82 x DRD
Slovakia	4	74.4	23.1	SZR, TDI	POU = 21.4 - 16.4 x SZR - 0.02 x TDI
Slovakia	4	99.5	98.6	DRD, TDI	POU = 12.3 - 0.58 x DRD - 0.01 x TDI
Spain	4	99.5	98.4	DLO, SZR	POU = 1.68 + 0.24 x DLO - 0.42 x SZR