



E.M.C.D.D.A.

EMCDDA SCIENTIFIC REPORT

European Network to Develop Policy Relevant Models and Socio-Economic Analyses of Drug Use, Consequences and Interventions

Final report: Part 2 – Prevalence of problem drug use at the national level

EMCDDA / 2002

EMCDDA SCIENTIFIC REPORT

**European Network to Develop Policy Relevant Models
and Socio-Economic Analyses of Drug Use, Conse-
quences and Interventions**

**Final Report: Part 2 –
Prevalence of problem drug use at the national level**

This report (Part 2 – National Prevalence) was prepared by:

Ludwig Kraus (work group coordinator) and Rita Augustin, IFT

Work Group members National Prevalence Estimation:

Rita Augustin, Catherine Comiskey, Ludwig Kraus, Petra Kümmler, Fabio Mariani, Carla Rossi, Alfred Uhl, Martin Frischer, Antonia Domingo-Salvany

Full Network Details:

Project Partners (project and work group coordinators):

Lucas Wiessing, EMCDDA (project coordinator), Gordon Hay, Univ. Glasgow, Carla Rossi, Lucilla Ravà, Univ. Rome 'Tor Vergata', Martin Frischer, Heath Heatlie, Univ. Keele, Hans Jager, Wien Limburg, RIVM, Christine Godfrey, Univ. York, Chloé Carpentier, Monika Blum, Kajsa Mickelsson, Richard Hartnoll, EMCDDA

The important input of all network participants and invited experts is fully acknowledged. For a list of network participants per working group and email contacts see Final Report Part 1, Annex A.

Other Network Participants and Invited Experts:

Erik van Ameijden, Fernando Antoñanzas, Ana Maria Bargagli, Massimiliano Bultrini, Marcel Buster, Maria Fe Caces, Maria Grazia Calvani, John Carnavale, Yoon Choi, Gloria Crispino O'Connell, Ken Field, Gerald Foster, Maria Gannon, David Goldberg, Peter Hanisch, Toon van der Heijden, Simon Heisterkamp, Matthew Hickman, Neil Hunt, Claude Jeanrenaud, Pierre Kopp, Mirjam Kretzschmar, Marita van de Laar, Nacer Lalam, Linda Nicholls, Alojz Nociar, Deborah Olzewski, Alessandra Nardi, Laetitia Paoli, Päivi Partanen, Paulo Penna, Harold Pollack, Maarten Postma, Thierry Poynard, Jorge Ribeiro, Francis Sartor, Janusz Sieroslawski, Ronald Simeone, Filip Smit, Juan Tecco, Alberto Teixeira, Jaap Toet, Gernot Tragler, Giovanni Trovato, Julián Vicente, Katalin Veress, Denise Walckiers, Robert Welte, Ardine de Wit, John Wong, Tomas Zabransky, Terry Zobeck, Brigitta Zuiderma-van Gerwen.

Project funded by the European Commission, DG Research, Targeted Socio-Economic Research (TSER). Project n^o: ERB 4141 PL 980030, Contract n^o: SOE2-CT98-3075

Starting date: 1st December 1998

Duration: 36 months

Date of issue of this report: 31st January 2002

© European Monitoring Centre for Drugs and Drug Addiction, 2002

Quotation is authorised providing the source is acknowledged.

European Monitoring Centre for Drugs and Drug Addiction

Rua da Cruz de Santa Apolónia 23-25

PT - 1149-045 Lisboa

Portugal

Tel: + 351 21 811 30 00

Fax: + 351 21 813 17 11

e-mail: info@emcdda.org

<http://www.emcdda.org>

Contents

Final Report Part 2: National Prevalence

1	Executive summary	5
2	Background and objectives	9
3	Description of the method	11
	3.1 Data Requirements.....	11
	3.2 Application.....	11
4	Data Sets.....	13
	4.1 Italian Data.....	13
	4.2 UK Data	20
	4.3 Western German Data	23
5	Data problems and analysis of their impact.....	27
6	Sensitivity analysis.....	31
	6.1 Violations of the requirements of the anchor points.....	31
	6.1.1 Effect of the number of anchor points and of the use of anchor points with high prevalence.....	32
	6.1.2 Only a part of a region is used as anchor point	34
	6.2 Violations of the requirements of the indicators.....	37
	6.2.1 Selection of indicators	37
	6.2.2 Systematic biases.....	47
	6.2.3 Indicator not reported by area of residence.....	51
	6.2.4 Influence of the range of the age group.....	52
7	Comparison of variants of the multivariate indicator method.....	55
	7.1 Different variants applied to actual data sets.....	56
	7.2 Simulation.....	62
8	Cross-validation.....	67
	8.1 Austrian Data.....	68
	8.2 PCA with indicator rates	68
	8.2.1 Shape of the relationship between factor scores and anchor point estimates.....	68

PART 2 – NATIONAL PREVALENCE - 3

8.2.2	Prevalence estimates	69
8.3	PCA with indicator values	71
8.3.1	Shape of the relationship between factor scores and anchor point estimates.....	71
8.3.2	Prevalence estimates	71
9	Dissemination and/or exploitation of results	73
	References.....	75
	Appendix.....	77
	SPSS-Syntax of the variant "PCA with original data".	78
	SPSS-Syntax of the variant "PCA per 100,000".	81
	SPSS-Syntax of the variant "PCA with ranks of the original data"	84
	SPSS-Syntax of the variant "PCA with ranks per 100,000".	87
	SPSS-Syntax of the variant "Correlation with original data".	90
	SPSS-Syntax of the variant "Correlation per 100,000".	94

For general overview or results of other working groups see:

Final Report Part 1	General Overview
Final Report Part 3	Work group 1b – Local Level Prevalence Estimation
Final Report Part 4	Work group 2a – Modelling Time trends and Incidence
Final Report Part 5	Work group 2b – Modelling Geographic Spread with Geographic Information Systems (GIS)
Final Report Part 6	Work group 3a – Modelling Costs and Cost-effectiveness of Interventions
Final Report Part 7	Work group 3b – Modelling Drug Markets and Policy options

1 Executive summary

The basic objective of the project work was to explore and to develop the multivariate indicator method. The method introduced by Person, Retka and Woodward (1977, 1978) and modified by Mariani (1999) estimates drug use by combining several population-standardized indicators directly corresponding to problematic drug use. With the use of principal component analysis, the complex information of the number of variables is reduced by extracting one single latent variable that is assumed to underlie all drug-related indicators, and that explains as much as possible of the variance of the original indicators. In a second step, the factor is used in a linear regression model with population-standardized prevalence estimates for at least two regions (the so-called anchor points). The linear regression results in population-standardized regional prevalence estimates. These are then used to calculate the national prevalence estimate.

Additionally, some variants of the method have emerged that differ in the way of transforming the indicator values (e.g. taking the logs, ranking, using the original values instead of the population-standardized ones) as well as in the method of reducing the information (principal component analysis, based on correlation matrix, summing up). Some of these variants were applied to existing data sets. Moreover, a cross-validation was conducted with an Austrian data set.

In the following, the results of these analyses are summarized:

1. At least three anchor points should be available, that should be from both sides of the continuum from low prevalence regions to high prevalence regions. The more anchor points are available, the more stable the method becomes towards other variations (such as choice of indicators, data weaknesses). **Implication:** Small scale studies are needed to provide a variety of independently obtained estimates. These studies should not be limited to areas with great drug problems, but also to areas with an assumed low prevalence.
2. The choice of indicators influences the model as well, however, this concerns mainly the rank of the regional prevalence estimates. **Implication:** Data collection should be organised rationally providing data collection and coding procedures that are comparable between the administrative regions. The choice of the drug-related indicators utilised for the study, however, is not yet final.
3. The method is relatively robust towards systematic biases of the indicators, e.g. the use of event-based data instead of person-based data in some or all regions, the inclusion of previous drug users or report not by area of residence. **Implication:** The method can be applied in spite of systematic biases.
4. The choice of the set of indicators should be theoretically based. Drug-related indicators representing consequences of problem drug use as e.g. treatment admissions or number of offences, cannot be easily replaced by social indicators. Aspects, such as face validity and basic assumptions, such as a monotonous relationship between drug prevalence and indicators should not be violated. **Implications:** Data on consequences of problem drug use should be made more easily available. If more indicators should be utilised, there should be empirical evidence that the indicator is drug-related.

5. Different variants of the method may result in a wide range of estimates. **Implication:** Different variants should be applied. In the case of rather different estimates it should be tried to find an explanation for the differences. At present, no recommendation for a certain variant can be given. The properties of the variants need further exploration.
6. As indicators are often not broken down by age group the choice of the age group is rather arbitrary. The choice of different age groups results in nearly the same regional and national prevalence estimates. **Implication:** To get prevalence estimates for the age groups recommended by EMCDDA a breakdown of the indicators and the anchor point estimates by age group is needed.
7. Overall the method seems to be appropriate for national prevalence estimation, but not for regional prevalence estimation. The choices of different sets of anchor points or indicators seem to effect more the regional prevalence rates than the national ones. In the sensitivity analysis and the cross-validation with capture-recapture estimates it turned out that changes of anchor points or indicators lead to high variations of the regional estimates – even if the national estimates are close to each other. **Implication:** Do not rely upon regional prevalence estimates obtained by the multivariate indicator method – especially if the regions are no anchor points.

Conclusions

From the effects above can be concluded, that the method works. The choice of the anchor point is crucial for the method but also the indicators should be selected carefully. The method is rather robust towards data flaws of the indicators, but it seems to be important that the indicators are consequences of problem drug use. However, there are still some properties of the method that could not be studied with the available data sets, such as the effects of anchor points estimates derived by different estimation methods and with different target groups or the effect of drug-related indicators not matching exactly to the target group of the anchor point estimates. It seems nearly impossible to analyse the latter problem as in practice no set of indicators will fit exactly to a the same, well-defined target group.

The influence of different methods for the anchor point estimates could, however, be analysed if at least two prevalence estimates derived with different estimations methods and/or different target groups were available for at least one of the anchor points. Even if the target groups are the same one method may be superior to the others, maybe due to obsolete multipliers or coverage errors.

Furthermore, at present we are unable to recommend the application of a certain variant of the multivariate indicator method. To create recommendations it would be necessary to apply the different variants of the method to many appropriate data sets, to compare the results and to conduct sensitivity analyses. Because of the high correlation between indicators it was impossible to apply the correlation variants to the Austrian data set whereas the German data set is inappropriate since all anchor points are high prevalence regions. Unless enough appropriate data sets were unavailable simulation studies could be conducted. To enable the simulation of realistic situations, profound examination of the distribution properties of commonly used indicators in many empirical data sets is necessary.

2 Background and objectives

Mostly, national prevalence of addiction is estimated by benchmark multiplier methods where the benchmark is obtained from a national data base and the multiplier is taken from a small scale study or an expert rating. The multivariate indicator method, however, introduced by Person, Retka and Woodward (1977, 1978) and modified by Mariani (1999) estimates drug use by combining several indicators directly corresponding to problematic drug use and regional prevalence estimates. The multivariate indicator method was agreed upon to be the most promising procedure for national prevalence estimation in the pilot-project on national prevalence estimation, as moreover, application of the multivariate indicator method is cheaper than nationwide capture-recapture estimation, which also combines different perspectives of the “drug problem”, e.g. the legal perspective, the medical/health perspective and the social perspective (EMCDDA, 2000a).

The method was first introduced by Person, Retka & Woodward^{1 2} and used to estimate the extent of problematic heroin use in the USA. A single latent variable is assumed to underlay the drug-related indicators, which can be extracted by principal component analysis. Generally, the indicator values are converted into rates per 100,000 and standardised. In a second step least squares regression is used to obtain the relationship between the prevalence rates of problem drug use (anchor points) and the values of the main factor, which explains most of the variation of the original variables. The linear regression allows the estimation of problem drug use in regions where only drug-related indicators are available. Independent prevalence estimates of at least two regions (anchor points) are necessary, possibly from regions with a low and a high prevalence rate respectively. The sum of all regional prevalence estimates yields the national prevalence estimate.

The aim of this project is the exploration of the properties of the multivariate indicator method. This includes the analysis of

- the impact of the anchor points: How do the estimates change if fewer anchor points are available? Is it possible to use local prevalence estimates as anchor points if not enough regional anchor points are available? Is it possible to use only high prevalence regions as anchor points?
- the impact of the indicators: To which extent are the estimates affected by systematic biases of the indicators? Which set of indicators should be selected? Does it matter if not all indicators are available? Can indicators presenting consequences of problem drug use be replaced with social indicators?

3 Description of the method

3.1 Data Requirements

The application of the multivariate indicator method requires a breakdown of national states by regions or provinces and data on problem drug use (indicators), which must be available for each of the regions and refer to the same time period. These indicator variables reflect the perspective of different societal systems (law enforcement system, health system..)

Examples of law enforcement data are:

- Data on seizures of controlled drugs
- Data on prices of illegal psychotropic substances
- Number of offenders against drug laws
- Number of convicted persons because of offences against drug laws
- Drug arrests
- Registered drug users

Examples of observations of the health system are:

- Cases of AIDS-infection related to intravenous use of psychotropic substances
- Number of drug-related deaths
- Emergency room drug abuse episodes
- Hospital based drug-related discharges
- Drug-related visits of general practitioners
- Drug abuse treatment admissions
- Drug-related ICD-9 or ICD-10 diagnoses and diagnostic related groups
- Number of methadone treatment admissions

Here the method is described for the case of five indicators, denoted by A, B, C, D and E. Additionally to the indicators, the population size F of the age group at risk in each region as well as independently obtained prevalence estimates G for at least two regions (the so-called anchor points) are needed. A_r denotes the value of the indicator A in region r, whereas A denotes the vector of regional indicator values. The same holds for B, C, etc. An SPSS syntax file is given in the appendix.

3.2 Application

For each of the indicators, for each anchor point estimate and for each region r the rate per 100,000 inhabitants is calculated by

$$A_{F,r}=A_r*100,000/F_r, \dots, E_{F,r}=E_r*100,000/F_r, G_{F,r}=G_r*100,000/F_r$$

The population standardized variables A_F, \dots, G_F have to be standardized (i.e., the difference between the regional value and the mean of all regions has to be divided by the standard deviation). With the use of principal component analysis, the complex information of the standardised indicators is reduced by extracting one single latent variable. In a second step, the unrotated first principal component (factor) is linked to G_F by linear regression with G_F as dependent variable and the coefficients of the first principal component as independent variable. The linear regression results in regression coefficients that enable the estimation of prevalence rates per 100,000 inhabitants for each region. Finally, these estimates have to be transformed to prevalence estimates for the regions. Summation of the regional prevalence estimates yields the national prevalence estimate.

Note, that we adopted the point of view of Person et al. (1977), who introduced the first principal component as a simple, one-dimensional indicator of problem drug use prevalence. Sartor & Walkiers (2001), however, point out that the number of principal components should be determined by the amount of variance that is explained. The variance of the chosen number of components should explain at least 75-80 percent of the total variance. Here a problem of interpretation emerges: What is the content of the second and further principal components? Can prevalence be seen as a multi-dimensional concept? If the first principal component explains only a small part of the total variance of the indicators this may reflect the lack of suitability of the indicators and it may be adequate to select a different set of indicators.

4 Data Sets

For the sensitivity analysis, data sets from three different countries have been used, that are described in more detail in the following sections. Furthermore, the prevalence estimates using the multivariate indicator method are presented.

4.1 Italian Data

For Italy data from 1995 and from 1996 are available (table 1 and table 2). While for 1995 the size of the 15-39-year-old population is given, the population size of the 1996 data refer to the 15-54-year-old population. In both years the indicators offences against drug laws, drug-related deaths, clients in treatment, AIDS related to IDU, and convictions of imprisoned addicts were collected.

Table 3 shows the rank order of the population standardized indicators. As can be seen from the ranks of the population standardized indicators, in 1995 all anchor points were regions with medium to high prevalence rates per 100,000 inhabitants. In 1996, however, both anchor points with low prevalence rates (Basilicata and Molise) and anchor points with high prevalence rates (Liguria and Valle d'Aosta) are employed. The ranks for 1995 and 1996 differ in 17 of the 100 pairs (20 regions multiplied by 5 indicators) by 5 ranks or more. Most of these big differences are found with indicator A (7 regions) and B (6 regions). Altogether, in 12 of the 20 regions these big differences are found with at least one indicator. Most of them are found in Trentino, where the ranks of the population standardized indicators A, B, and D differ by 5 ranks or more.

Figures 1 and 2 show the linear regressions with the prevalence rates of the anchor points as dependent variable and the factor scores as independent variables for 1995 and 1996 respectively. Regarding the anchor points, the position fits to the conclusions drawn from table 3: In 1995, the anchor points Lazio and Lombardia exhibit comparatively high factor scores, while Sardegna has a medium factor score. In 1996, the anchor points are the regions with the two highest (Liguria and Valle d'Aosta) and the two lowest (Basilicata and Molise) factor scores.

Table 1: Parameters and anchor points for the multivariate indicator method for Italy, 1995

Regions	Population 15-39 years	A	B	C	D	E	G anchor points
Abruzzo	458,886	384	22	2729	22	384	
Basilicata	231,425	130	6	796	9	70	
Calabria	798,753	516	8	2637	39	167	
Campania	2,302,156	1809	109	10002	167	1260	
E. Romagna	1,356,948	1494	88	8326	402	987	
Friuli	414,965	413	15	2338	32	214	
Lazio	1,963,614	2517	112	9691	514	1996	53,733
Liguria	544,707	824	47	3586	159	387	
Lombardia	3,245,393	3625	185	19608	1004	1943	42,297
Marche	501,843	463	13	2378	63	199	
Molise	118,771	61	3	346	2	60	
Piemonte	1,508,778	2191	88	12564	281	1445	
Puglia	1,607,028	1016	42	10543	150	892	
Sardegna	671,185	484	16	5238	130	504	13,618
Sicilia	1,933,259	994	34	6311	170	694	
Toscana	1,209,749	1187	64	10034	245	740	
Trentino	348,195	457	22	1447	53	135	
Umbria	278,902	236	14	1877	34	194	
Valle d'Aosta	43,525	49	0	289	5	49	
Veneto	1,684,716	1202	92	10222	222	711	
Total	4408,605	20052	980	120962	3703	13031	

A Offences against drug laws

B Drug-related deaths

C Clients in treatment

D AIDS related to IDU

E Convictions of imprisoned addicts

G Estimated values of regional IDU population, independently obtained

Table 2: Parameters and anchor points for the multivariate indicator method for Italy, 1996

Regions	Population 15-54 years	A	B	C	D	E	G anchor points
Abruzzo	694026	769	21	3145	20	503	
Basilicata	333706	262	6	982	11	98	1267
Calabria	1156588	1190	27	3491	37	204	
Campania	3325084	3053	162	11625	108	1522	
Emilia Romagna	2138085	2421	123	8943	300	1065	
Friuli V.G.	653802	600	18	2264	26	288	
Lazio	2989586	3813	204	9864	361	1594	
Liguria	857029	1394	108	2244	178	761	9127
Lombardia	5159928	3958	261	20666	941	1931	
Marche	777232	772	30	3869	44	238	
Molise	176709	108	3	535	3	72	1185
Piemonte	2,368,358	2801	161	13248	194	1395	
Puglia	2347101	2318	81	11539	161	1044	
Sardegna	989869	606	37	5439	115	772	
Sicilia	2839180	2573	60	7842	78	844	
Toscana	1907134	2674	78	9611	144	1039	
Trentino A.A.	523936	554	18	987	29	138	
Umbria	442168	536	18	2811	27	251	
Valle d'Aosta	67860	109	8	341	4	87	989
Veneto	2538118	2474	97	10438	156	849	
Total	32,315,499	32985	1521	129884	2937	14695	

A	Offences against drug laws
B	Drug-related deaths
C	Clients in treatment
D	AIDS related to IDU
E	Convictions of imprisoned addicts
G	Estimated values of regional IDU population, independently obtained

Table 3: Rank order of the population standardised indicators and anchor point estimates for Italy, 1995 and 1996

Regions	A		B		C		D		E		G anchor points	
	1995	1996	1995	1996	1995	1996	1995	1996	1995	1996	1995	1996
Abruzzo	9	13	11	6	10	13	3	3	17	17		
Basilicata	3	4	7	2	4	4	2	6	2	3		1
Calabria	5	11	2	4	3	5	4	4	1	1		
Campania	8	7	10	14	6	9	5	5	9	11		
Emilia Romagna	14	14	19	16	13	12	19	18	15	12		
Friuli V.G.	13	6	9	5	9	8	6	7	8	9		
Lazio	17	17	16	18	8	7	17	17	19	13	3	
Liguria	20	20	20	20	15	2	18	20	14	19		3
Lombardia	15	3	15	15	11	10	20	19	11	7	1	
Marche	11	10	6	11	7	15	11	9	5	5		
Molise	1	1	5	1	1	6	1	1	7	8		2
Piemonte	19	15	17	17	20	19	14	15	18	16		
Puglia	4	9	8	8	14	14	8	13	10	10		
Sardegna	7	2	4	9	18	18	15	16	16	18	2	
Sicilia	2	5	3	3	2	3	7	2	3	4		
Toscana	12	18	13	13	19	17	16	14	12	14		
Trentino A.A.	18	12	18	7	5	1	13	8	4	2		
Umbria	10	16	12	12	17	20	10	12	13	15		
Valle d'Aosta	16	19	1	19	16	16	9	10	20	20		4
Veneto	6	8	14	10	12	11	12	11	6	6		

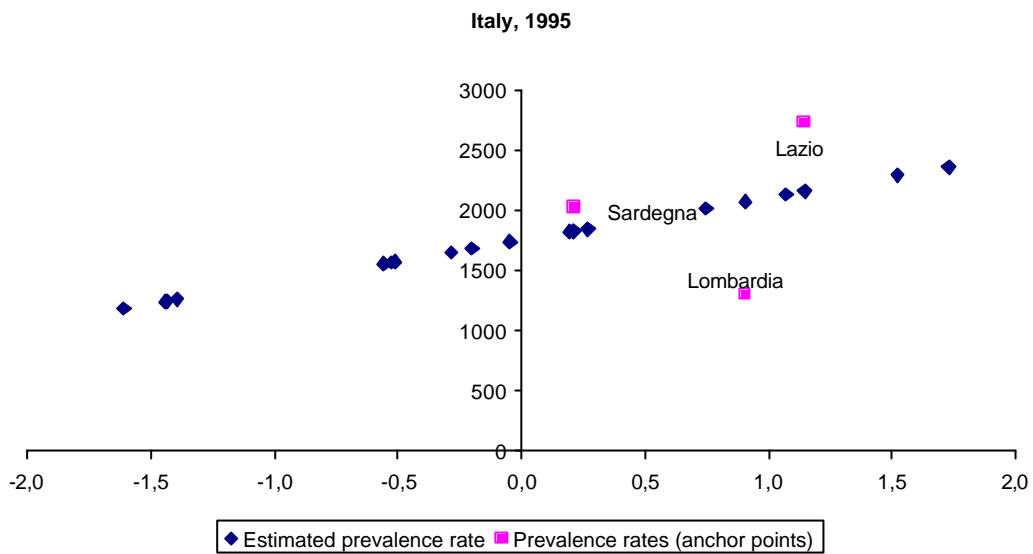


Figure 1: Regression line indicating relationship between factor scores and population standardized anchor point estimates, Italy 1995

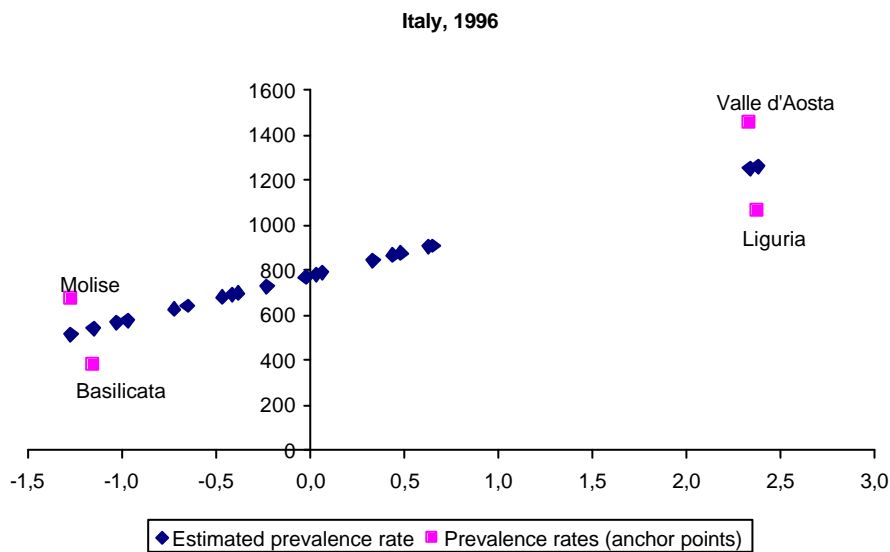


Figure 2: Regression line indicating relationship between factor scores and population standardized anchor point estimates, Italy 1996

PART 2 – NATIONAL PREVALENCE - 18

The regional prevalence estimates are depicted in figures 3 and 4. As can be seen immediately, the 1995 estimates clearly exceed the 1996 estimates. In most of the regions the 1996 prevalence estimate amounts to about two thirds of the 1995 estimate, resulting in a national estimate of 386,112 for 1995 (15-39 years) and 248,720 for 1996 (15-54 years).

In figure 5 the 1995 factor scores and the 1996 factor scores are compared. The scatterplot shows a nearly linear relationship between the factor scores of these two years. Outliers are Valle d'Aosta and Liguria, which were used as anchor points in the 1996 data set. Compared to the other points, the 1996 factor scores of Valle d'Aosta and Liguria are too high. As can be seen from figure 2, if the factor scores of these two anchor points were smaller the regression line would be steeper, resulting in higher regional prevalence estimates for 1996.

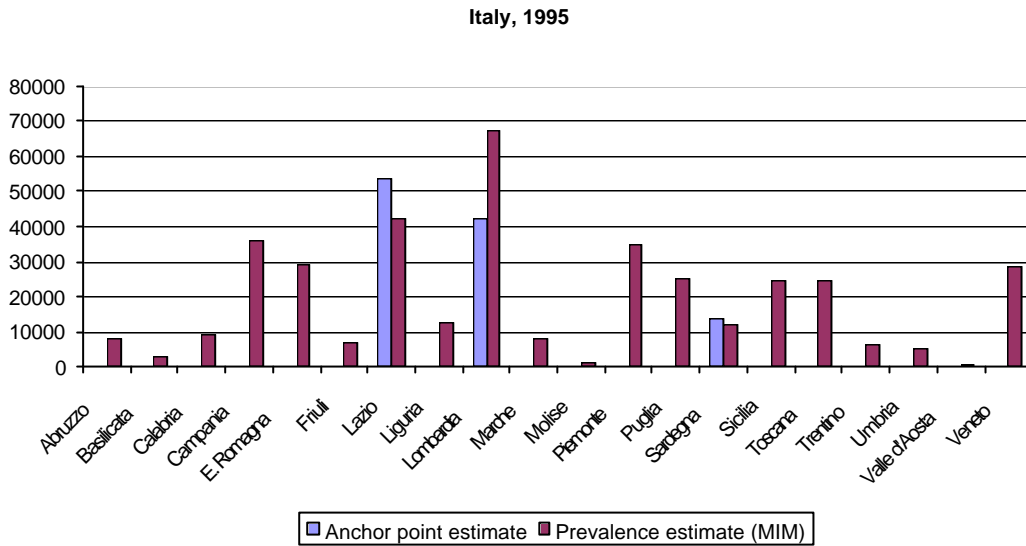


Figure 3: Regional prevalence estimates, Italy 1995

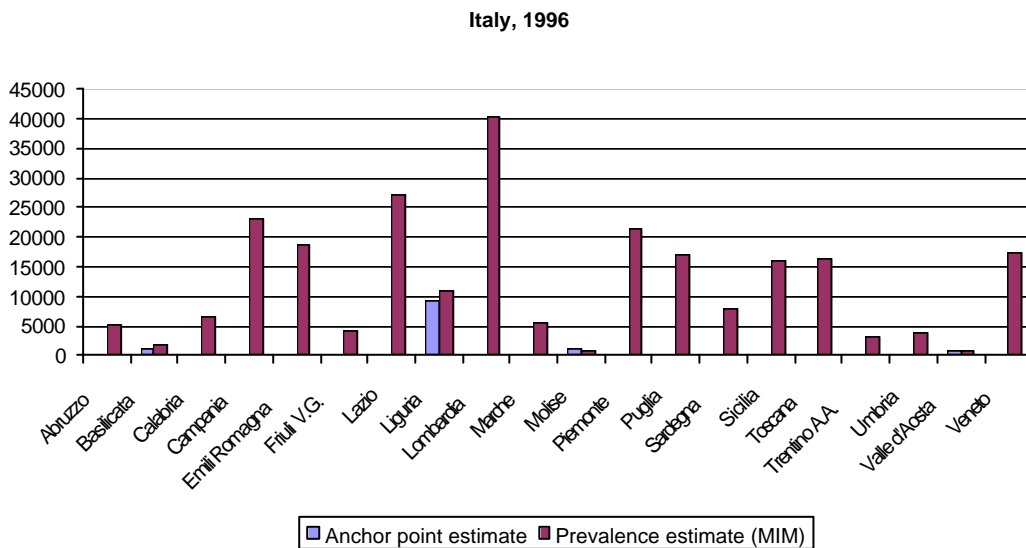


Figure 4: Regional prevalence estimates, Italy 1996

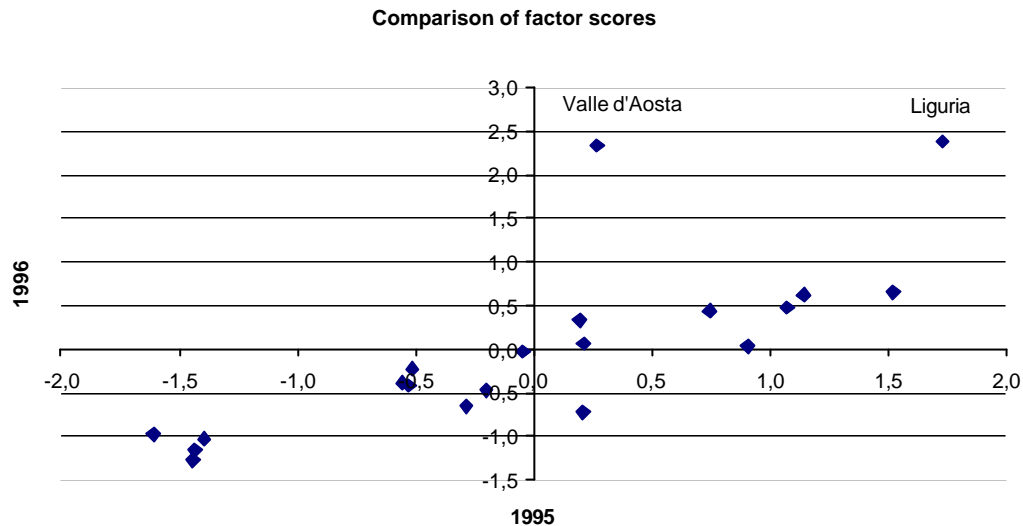


Figure 5: Scatterplot of factor scores, Italy 1995 and 1996

This comparison of results from two adjacent years from the same country demonstrates the high sensitivity of the multivariate indicator method towards indicators and prevalence estimates of the anchor points.

4.2 UK Data

For the UK the indicators convictions for drug offences, seizure of controlled drugs, people receiving treatment for drug misuse, cases of AIDS related to IDU and drug-related deaths were employed. Except for drug-related deaths and treatment data the indicators refer to the year 1996. In the case of England and Wales the number of clients in treatment between October 1996 and March 1997 is given, while for Scotland the number of clients in treatment refers to the time period April 1995 to March 1996. For all regions, the number of drug-related deaths in 1995 is shown. Note, that seizures and convictions are event-based, while the other indicators are person-based. The anchor point estimates were obtained through various techniques and are for different time periods (Frischer et al., 2001). No specific age group was selected, the given population sizes are the sizes of the whole population. The data set given in table 4 was taken from the final report of the EMCDDA project "Study to obtain Comparable National Estimates of Problem Drug Use Prevalence for all EU Member States" (EMCDDA, 2000a) and differs slightly from the data set in Frischer et al. (2001).

Table 4: Parameters and anchor points for the multivariate indicator method for the UK

Regions	Population	A	B	C	D	E	G anchor points
Northern and Yorkshire	6,600,626	11356	13285	9722	37	344	
Trent	4,606,495	6451	7010	3580	67	207	
Anglia and Oxford	4,521,912	3761	4183	3762	79	216	
North Thames	7,190,479	17696	21168	7842	334	352	44410
South Thames	6,579,403	13987	16530	7774	122	346	38140
South West	6,131,705	10600	12717	5890	60	311	
West Midlands	5,150,246	7125	5398	4322	26	193	13130
North West	6,274,338	12557	11804	8958	63	402	
Wales	2,835,073	6110	5870	2282	14	139	8357
Scotland	5,120,000	3008	13452	8614	687	267	38000
Total	59,283	92651	111417	62746	1489	2777	

- A Convictions for drug offences
- B Seizure of controlled drugs
- C People receiving treatment for drug misuse
- D Cases of AIDS related to IDU
- E Drug-related deaths
- G Independently obtained estimated number of problematic drug users

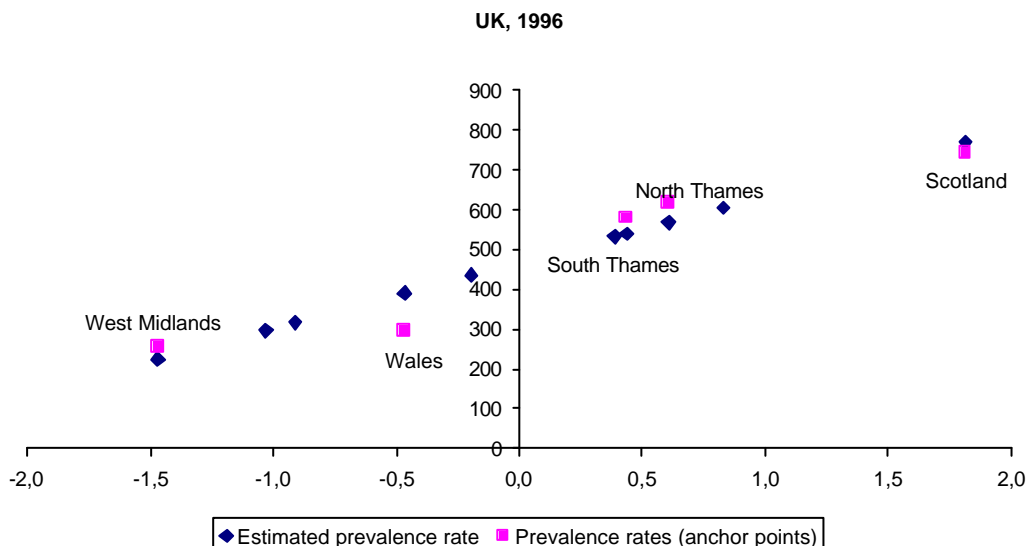


Figure 6: Regression lines indicating relationship between factor scores and population standardized anchor point estimates, UK 1996

PART 2 – NATIONAL PREVALENCE - 22

As can be seen from figure 6 both the regions with the lowest (West Midlands) and the highest (Scotland) factor scores are anchor points. The anchor point with the second highest factor score is North Thames, followed by South Thames, and Wales.

The regional prevalence estimates are depicted in figure 7. The highest prevalence was found in North Thames, the lowest in Wales. The national estimate is 265,944.

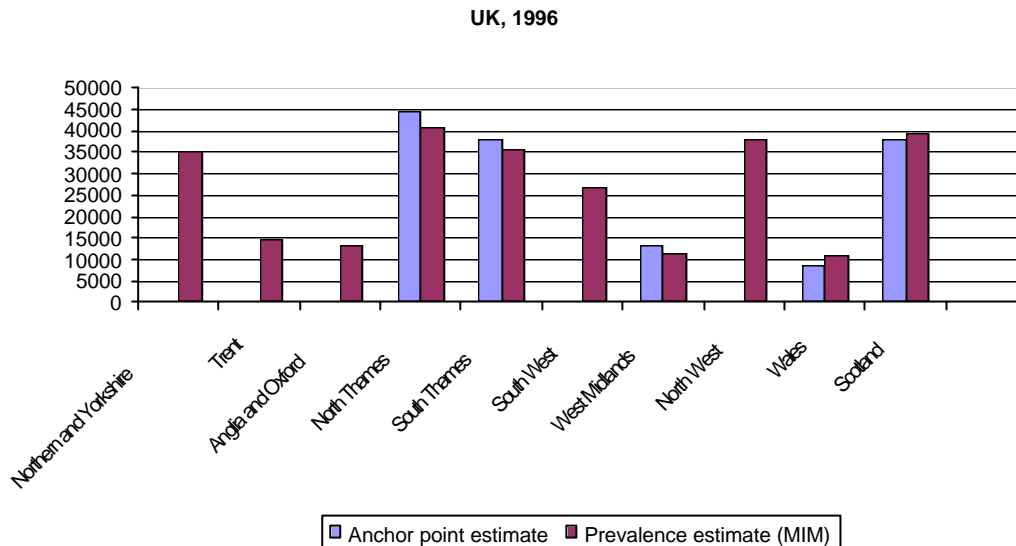


Figure 7: Regional prevalence estimates, UK 1996

4.3 Western German Data

In Germany, anchor point estimates for 1995 are available. These are, however, not results of prevalence studies but extrapolations of the prevalence rate of 0,325% i.v. drug users in western German cities with at least 500,000 inhabitants (Kirschner & Kunert, 1997). This figure was obtained through a nation-wide survey among medical doctors. As three of the German regions are also metropolis, anchor point estimates were obtained by multiplying the size of the whole population of these regions with 0,325%. In the case of Berlin, the prevalence estimate refers only to west Berlin.

Indicators are offences against drug laws, drug-related deaths, clients in treatment, AIDS related to i.v. drug use, and convictions of imprisoned addicts. Obviously, only the AIDS indicator is related to intravenous drug use, while the other indicators may include also drug users with a different way of application. Offences against drug laws and convictions of imprisoned addicts are event-based, not person-based. Moreover, treatment data are obviously imprecise as in all regions the number of clients in treatment is a multiple of 100.

These indicators are available for all German regions. As in East Germany illicit drugs were nearly unavailable before the re-unification in 1990 indicators as e.g. the number of drug-related deaths seem to be inappropriate for East Germany because of the long latency time. Thus, estimation is restricted to Western Germany. The data are given in table 5.

Table 5: Parameters and anchor points for the multivariate indicator method in Western Germany

Western German regions, 1995	Population 15-54 years	A	B	C	D	E	G anchor points
Baden-Württemberg	5774803	13225	255	17500	317	1084	
Bayern	6694822	9538	224	12500	307	581	
Berlin	2047829	5507	93	7500	441	599	7073
Bremen	376691	2689	51	3000	51	145	1798
Hamburg	973412	6827	141	8500	117	674	5444
Hessen	3384030	7825	166	9000	256	370	
Niedersachsen	4262229	8020	99	9000	91	1017	
Nordrhein-Westfalen	9812031	26759	380	31000	424	2442	
Rheinland-Pfalz	2161042	3594	69	5500	63	97	
Saarland	588738	1043	25	1200	19	34	
Schleswig-Holstein	1501615	1585	53	3600	21	80	
Total	37577203	86612	1556	108300	2107	7123	

A Offences against drug laws
 B Drug-related deaths
 C Clients in treatment
 D AIDS related to IDU
 E Convictions of imprisoned addicts
 G Estimated values of regional IDU population, independently obtained

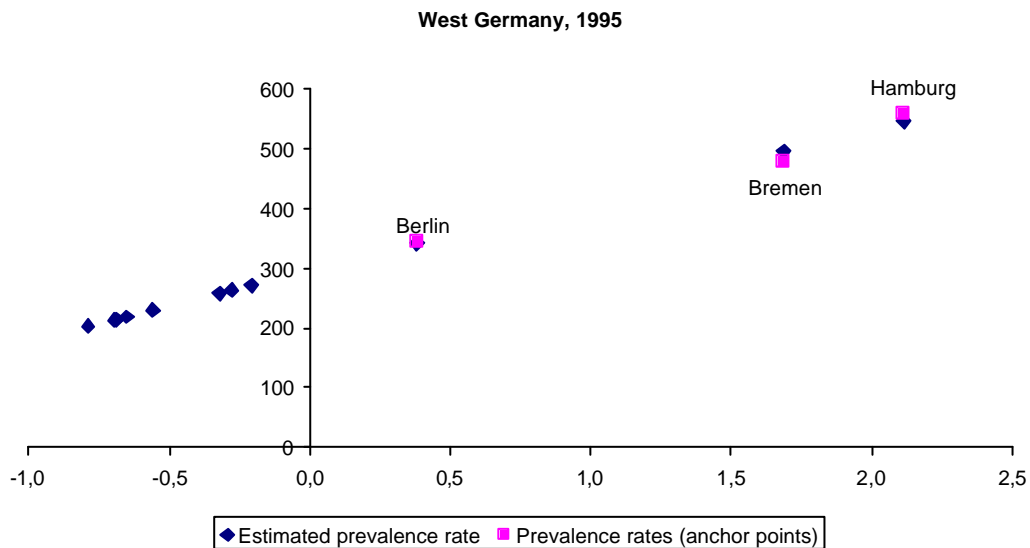


Figure 8: Regression lines indicating relationship between factor scores and population standardized anchor point estimates, Western Germany, 1995

As can be seen from figure 8 the regions with the three highest factor scores are anchor points. Since the number of inhabitants of these regions is low, the estimated number of i.v. drug users is low compared to the three biggest regions Nordrhein-Westfalen, Bayern, and Baden-Württemberg (figure 9). The national estimate is 97,834, which is about two thirds of the national estimate of 150,406 obtained in the nation-wide survey on medical doctors from which the anchor point estimates were taken (Kirschner & Kunert, 1997).

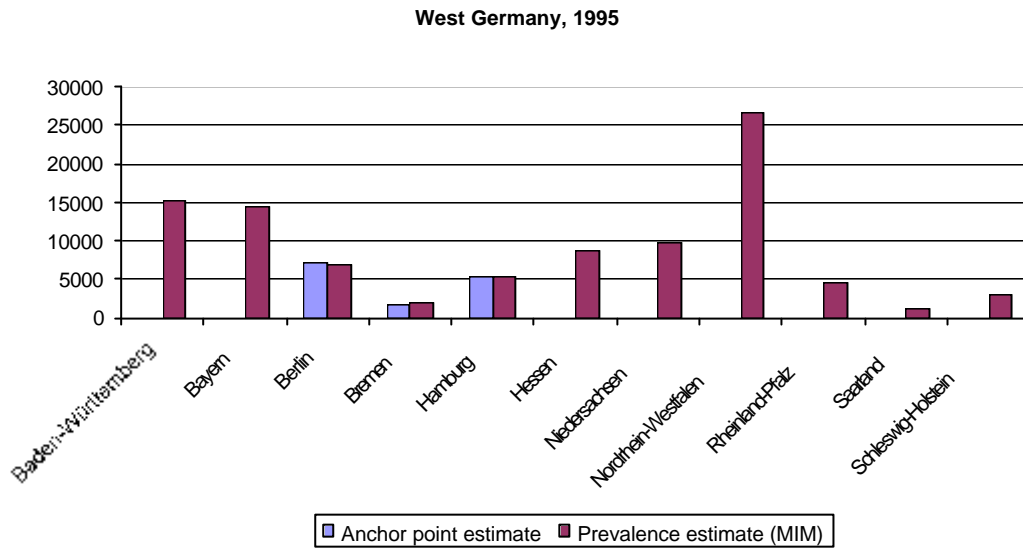


Figure 9: Regional prevalence estimates, Western Germany, 1995

5 Data problems and analysis of their impact

Each method underlies certain assumptions and requirements on data quality. In the case of the multivariate indicator method both the anchor points and the indicators should fulfil certain requirements.

These include in the case of the anchor points:

- There should be at least two anchor points.
- The target groups of all anchor point estimates should be the same.
- Anchor points should be taken both from low prevalent regions and from high prevalent regions.
- The anchor points should be regions employed in the PCA step.

The requirements on the indicators are:

- The indicators should be correlated to problem drug use. Ideally, there should be a linear relationship.
- The same decomposition in regions and the same time period should be used for all indicators.
- There should be no systematic bias. Systematic biases may be caused by the use of event-based data, by delays in data entry or by the use of time frames of different length.
- The variables should be reported by geographical area of residence.
- The indicators should refer to the age group that is employed in calculating the figures per 100,000 inhabitants.

Moreover, there are certain requirements on the relationship between anchor points and indicators:

- The indicators should fit to the target group of the anchor point estimates.
- The anchor point estimates and the indicators should refer to the same time period.
- There should be a linear relationship between factor scores derived from the indicator rates and the prevalence rates of the anchor points.

Apparently, these requirements are often not fulfilled in practice. Regarding the requirements on the anchor points, many countries do not have reliable prevalence estimates of two or more regions. Often only prevalence estimates of cities, but not of the surrounding regions are available. If problem drug use is concentrated heavily in these cities they may be used as anchor points. In most of the European countries, however, problem drug use is spreading in rural areas. Then the problem may be handled by dividing the region with the local prevalence estimate in two new regions – the city that serves as anchor point and the rest of the region. Both solutions, however, imply a further problem: If only cities are employed as anchor points the requirement that also low prevalent regions should be available as anchor points is not fulfilled as normally the prevalence

rates of cities exceed the prevalence rate of rural areas. The sole use of high prevalent regions may make the linear regression very “unstable” as the slope of the line is largely dictated by the nonsystematic component of the variation (Wickens, 1993). This is illustrated in figure 10. The anchor points, which are shaded, are adjacent. The regression line through the anchor points deviates heavily from the regression line we would have obtained if we were able to replace one anchor point by another region with a lower prevalence. Obviously, it is easier to get high prevalent anchor points than low prevalent anchor points since scientific projects are more often conducted in regions where the drug problem has become apparent.

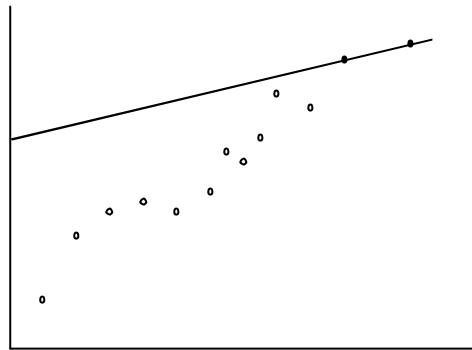


Figure 10: Illustration of the possible effects of too similar anchor points

Evidently, increasing the number of anchor points also makes the regression more stable. It may be reasonable to employ anchor point estimates obtained with different methods and targeting different sub-groups of problem drug users to get a high number of anchor points. This was done in the case of the UK data set, where each second region is an anchor point. As the target group of the anchor point estimates determines the target group of the national prevalence estimate the use of different target groups for different anchor points brings about the problem that it is unclear what we are estimating! If the overlap between the different target groups is large the uncertainty regarding the target group of the national prevalence estimate should be tolerated. In practice, however, the extent of the overlap is not known as additionally to the problem of different target groups there is a variety of different other problems as e.g. doubt on the validity of multipliers, underreporting, double counting and so on.

The impact of weaknesses of the anchor points was analysed with the UK data set. By employing the two anchor points with the highest prevalence (Scotland and North Thames) as anchor points we investigated the impact of the lack of anchor points with a low prevalence. Furthermore, we calculated different models with different anchor points to get hints on the minimum number of anchor points that are required to get a good national prevalence estimate. We analysed the problem that a good prevalence estimate is only available for a part of a region by replacing the prevalence estimate for Scotland with the prevalence estimate for Strathclyde and decomposing Scotland in Strathclyde and the rest of Scotland. We were not able to analyse the impact of different target groups of the anchor point estimates as we do not have a data set with two prevalence estimates for the same anchor point.

The different indicators highlight different aspects of the drug problem. No indicator is supposed to measure prevalence. The indicators are, however, indicative of whether problem drug use increases or decreases (Person et al., 1977). By applying principal component analysis a common factor is extracted which is assumed to be proportional to prevalence of problem drug use. As principal component analyses underlies the assumption of a linear relationship between observable variables and the principal components there should be a linear relationship between indicators of problem drug use and the unknown prevalence. In practice, this assumption is, however, violated as e.g. in the case of treatment data the number of drug users in treatment may be restricted by the capacity of treatment services. With regard to seizures, increasing prevalence and thus a increasing number of seizures will lead to more police enforcement and therefore contribute to a further increase in seizures. Person et al. (1976, 1977) suggest the use of rank-ordered indicator values as they assume that the relationship between those and the unknown drug prevalence deviates less from a straight line than the relationship between the original indicators and drug prevalence. We compare results obtained both by employing the original data and the rank-ordered data in chapter 7.

Due to a lack of available drug-related indicators the Dutch work group of the project “Study to obtain Comparable National Estimates of Problem Drug Use Prevalence for all EU Member States” (EMCDDA, 2000a) used an alternative model with social indicators. We re-analysed the German data with these indicators to examine if this alternative is reasonable in other countries without drug-related indicators.

Some of the other requirements are often met in practice. Many countries are decomposed in regions and all the authorities use the same decomposition of regions. As data often are published yearly the requirement of the same time frame for all indicators does not lead to any problems. Usually police statistics are not reported by area of residence but by area of report. If the regions are big enough and the big cities are located in the middle of the regions the violation of this requirement will not influence the regional and the national prevalence estimates. If, however, a metropolis is situated near the border to another region residents of the other region have a high probability of being seized in the city. The impact of this problem is analysed with the German data set where three regions are cities by lessening the offences in the cities and increasing this number in adjacent regions

With regard to police data a further problem emerges: Often not the number of offenders, but the offences is reported, i.e. the statistic is event-based, not person-based. As more than one offender can be involved in the same offence and one individual can commit more than one offence the relation between number of persons and number of events is not clear. In other cases, the reported figures often exceed the true figures. If e.g. treatment data reflect the number of treatments instead of the number of treated individuals the reported figures are too high as one individual can be treated more than once. Apart from systematic biases due to the use of event-based statistics, other systematic biases occur: If data entry is delayed the reported figures are smaller than the true ones. In the case of AIDS of i.v. drug users, a part of these individuals may have ceased injecting. Then the reported figures are too high.

We analysed systematic biases by multiplying the values of one indicator of about half of the German and British regions with 5. This factor was chosen arbitrarily and the approach is supposed to generate a situation worse than that in practice, as that large differences between reported and true figures are not expected. We refrained from changing all indicator values in this way as this would not change the factor score at all.

Often, indicators are not broken down by age group. Thus, the choice of the age group in the application of the multivariate indicator method is rather arbitrary. We analysed the impact of the age group by using two different age groups with the German data.

Apparently, the shape of the relationship between factor scores and anchor point estimates is of crucial importance. If it deviates substantially from a straight line, linear regression yields biased results and nonlinear regression should be applied. In practice, the low number of anchor points compared to the number of regions hinders the selection of the appropriate regression function. The shape of the relationship between factor scores and prevalence rates of the anchor point estimates is studied with an Austrian data set, where for each region a capture-recapture estimate was calculated in chapter 8. Though half of the regions are anchor points the UK data set is inappropriate for the analysis of this question as target groups, estimation methods, and referred time periods of the anchor point estimates differ.

Due to the lack of appropriate data we were not able to analyse the impact of the other violations of requirements on the relationship between anchor points and indicators. It seems, however, nearly impossible to study the impact of indicators not matching exactly to the target group of the anchor point estimates as in practice no indicator will fit exactly to a well-defined target group. As Frischer et al. (2001) point out, the total number of drug-related deaths may include many cases where the person was not a problem drug user (i.e. it may have been their first experience of drug use). Obviously, these coverage error is even more likely with police data. It is even harder to find a set of indicators that refer to – more or less – the same target group. Apart from a few exceptions, the mortality data are a subset of treatment data, which in turn are a subset of police data: The majority of drug overdose deaths had used drugs intravenously. Treatment data, however, cover intravenous drug users, but also problematic consumers using other routes of administration with a definitely smaller risk of fatality. Police data include an even larger population, since non problem drug users also may have been registered by the police. Moreover, the mortality data are – more or less – a subset of the HIV/AIDS data as the latter cover lifetime drug users that have used drugs intravenously and became infected with HIV while all the other indicators refer to a 12-month period.

6 Sensitivity analysis

The impact of different data problems was analysed as described in chapter 5. Section 6.1 presents the results concerning problems with anchor points. In section 6.2 the impact of problems with indicators on the prevalence estimate is examined. The question of the shape of the relationship between factor scores and anchor point estimates and thus the question if linear or nonlinear regression should be applied is postponed to chapter 8.

6.1 Violations of the requirements of the anchor points

Calculations with different anchor points were based on the British data, as five anchor points were available. Different models based on various combinations of anchor points were calculated and compared.

Figure 11 depicts the ranking of prevalence rates in the UK as it is obtained in the principal component step of the multivariate indicator method. Scotland is highest in prevalence rate, followed by North West, North Thames and South Thames. The region Northern and Yorkshire, which will be analysed in more detail, is rather similar in prevalence rate to South Thames. The lowest prevalence rates are found in Trent, Anglia and Oxford and in the West Midlands.

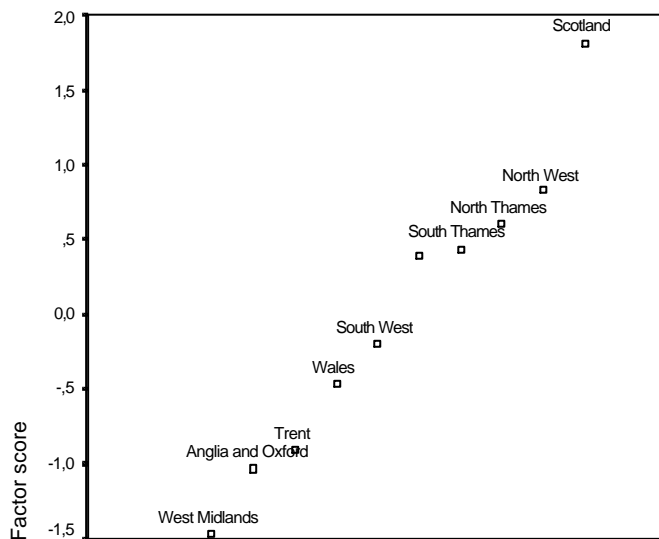


Figure 11: Ranking of prevalence rates in the UK

6.1.1 Effect of the number of anchor points and of the use of anchor points with high prevalence

Calculations with different anchor points were based on the British data, as five anchor points were available. Different models on various combinations of anchor points were calculated and compared to get hints on the minimum number of anchor points that are required to get a good national estimate. As in practice reliable prevalence estimates for regions with a high prevalence are more easily available the focus is on the effect of the exclusive use of anchor points with a high prevalence.

Table 6 shows the variation of the national prevalence estimates depending on the different models (calculations with different anchor points and different number of anchor points). The first column lists the utilised anchor points, the second the national prevalence and the third column shows the difference referring to the result of the original model.

Table 6: Effects of different anchor points: changes of national prevalence of the UK

Anchor Points	Prevalence UK	Difference
North Thames, South Thames, West Midland, Wales, Scotland	265,944	Reference
North Thames, Scotland	310,855	44,911
North Thames, South Thames	277,032	11,088
West Midland, Wales	174,420	-91,524
Wales, Scotland	222,921	-43,023
West Midland, Scotland	268,204	2,260
North Thames, West Midland, Scotland	277,858	11,914
North Thames, South Thames, Scotland	302,704	36,760
South Thames, West Midland, Wales	275,734	9,790
North Thames, South Thames, West Midland, Wales	272,799	6,855

The total prevalence estimates derived by the different calculations reached from around 200,000 problematic drug users to more than 300,000 – which is a difference of more than 30%. The original calculation including all anchor points gives an estimate of 266,000 problematic drug users. Calculations with only two anchor points result in the highest differences.

Table 7 shows the effects of the choice of different anchor points and the choice of a varying number of anchor points on the prevalence rates of the regions. Each column presents the results of the calculation of one model. Regions utilised as anchor points are marked with asterisk [*]. The columns are ordered by number of anchor points starting from the left side of the table with combinations of two anchor points to the right side of the table with five anchor points.

Table 7: Effects of different anchor points: prevalence rates of the UK and its regions

Region	Prevalence/100,000									
Northern & Yorkshire	595	569	329	462	531	549	583	545	550	569
Trent	461	278	277	207	338	353	437	350	310	278
Anglia & Oxford	449	251	272	184	320	335	424	333	288	251
North Thames	* 618	* 618	337	505	563	* 582	* 608	577	* 590	* 618
South Thames	600	* 580	331	472	538	556	* 589	* 552	* 559	* 580
South West	535	438	305	348	444	461	517	457	442	438
West Midlands	403	152	* 255	97	* 255	* 268	374	* 267	* 206	* 152
North West	641	668	346	549	596	616	633	611	631	668
Wales	507	378	* 295	* 295	404	420	487	417	* 392	* 378
Scotland	* 742	888	385	* 742	* 742	* 765	* 743	* 758	813	* 888

It can be seen that the prevalence rates in a row differ, i.e. they differ between the various models. However, the results become more stable with the use of at least three anchor points (compare the models with five, four anchor points and three anchor points). To illustrate this, in the following table 8 the prevalence rates for the region Northern & Yorkshire are presented. The first column shows the anchor points utilised, the second column the resulting prevalence rate and the third column presents the difference referring to the model with five anchor points.

Table 8: Effects of different anchor points for the region Northern & Yorkshire

Anchor Points	Northern & Yorkshire	
	Prevalence rate	Difference
North Thames, South Thames, West Midland, Wales, Scotland	532	Reference
North Thames, Scotland	595	63
North Thames, South Thames	595	37
West Midland, Wales	569	-203
Wales, Scotland	329	-70
West Midland, Scotland	462	-1
North Thames, West Midland, Scotland	531	17
North Thames, South Thames, Scotland	549	51
South Thames, West Midland, Wales	583	13
North Thames, South Thames, West Midland, Wales	545	18

The differences are biggest for calculations connecting only two anchor points in a linear regression model, but it seems also important which anchor points are used. The anchor points North Thames and South Thames result in a lower deviation than West Midland and Wales or Wales and Scotland. Another example is given in table 9, which presents the results for the region Trent. However, for Trent the model with the three anchor points North Thames, South Thames and Scotland results in the greatest variation, thus, not only the number of anchor points seems to be relevant, but also other effects, such as how representative the anchor points are.

Table 9: Effects of Anchor Points: Changes in Prevalence Rates for the Region Trent

Anchor Points	Trent	
	Prevalence rate	Difference
North Thames, South Thames, West Midland, Wales, Scotland	316	Reference
North Thames, Scotland	461	145
North Thames, South Thames	277	-38
West Midland, Wales	207	-39
Wales, Scotland	338	-109
West Midland, Scotland	353	22
North Thames, West Midland, Scotland	437	37
North Thames, South Thames, Scotland	350	121
South Thames, West Midland, Wales	310	34
North Thames, South Thames, West Midland, Wales	278	-6

The analysis showed that the total prevalence is highly dependent on the choice of anchor points as having been described by Person et al. (1976), as these anchor points give the actual span of prevalence between which the regions are spread. Therefore, at least three anchor points should be available, that should be from both sides of the continuum, i.e. from low prevalence regions to high prevalence regions.

6.1.2 Only a part of a region is used as anchor point

Often only prevalence estimates of cities, but not of the surrounding regions are available. This problem may be handled by dividing the region with the local prevalence estimate in two new regions – the city that serves as anchor point and the rest of the region. We analysed the problem that a good prevalence estimate is only available for a part of a region by replacing the prevalence estimate for Scotland with the prevalence estimate for Strathclyde and decomposing Scotland in Strathclyde and the rest of Scotland. We analysed two questions:

1. Do the results differ substantially if instead of the whole region (Scotland) only a part of it (Strathclyde) is employed as anchor point?
2. Can estimation be improved if a part of a region is used as an additional anchor point?

We compared both the results based on all five anchor points and the results based on the anchor points apart from Scotland with the results based on the anchor points Strathclyde, North Thames, South Thames, West Midlands, and Wales. As mostly fewer anchor points are available

and moreover the influence of a single anchor point on the national estimate tends to decrease with the number of anchor points we compared the results of three further sets of anchor points: 1) Wales and North Thames, i.e. the anchor points with the lowest and the second highest prevalence estimates per 100,000 inhabitants 2) Wales, North Thames, and Scotland, and 3) Wales, North Thames, and Strathclyde. The different national prevalence estimates are presented in table 10.

Table 10: Effects of Strathclyde as anchor point: changes of national prevalence of the UK

Anchor Points	Prevalence UK	Difference
North Thames, South Thames, West Midland, Wales, Scotland	265,944	Reference
North Thames, South Thames, West Midland, Wales, Strathclyde	270,215	4,271
North Thames, South Thames, West Midland, Wales	272,799	6,855
North Thames, Wales, Scotland	244,132	-21,812
North Thames, Wales, Strathclyde	261,037	-4,907
North Thames, Wales	255,803	-10,141

With regard to the first question, we found a negligible difference between the two results based on five anchor points and a larger difference if we compare the results employing three anchor points. Here, however, the estimate using Strathclyde is closer to the “original” estimate using North Thames, South Thames, West Midland, Wales, Scotland than the estimate using Scotland! With regard to the second question, including Strathclyde decreases the deviation from the reference value.

The difference is, however, negligible, if all anchor points except for Scotland are employed. Figure 13 reveals that in this case not only the national but also the regional estimates are rather close to each other, whereas including Strathclyde has a comparatively high impact on e.g. the estimates of West Midlands or Scotland in the case of the two anchor points North Thames and Wales (figure 12).

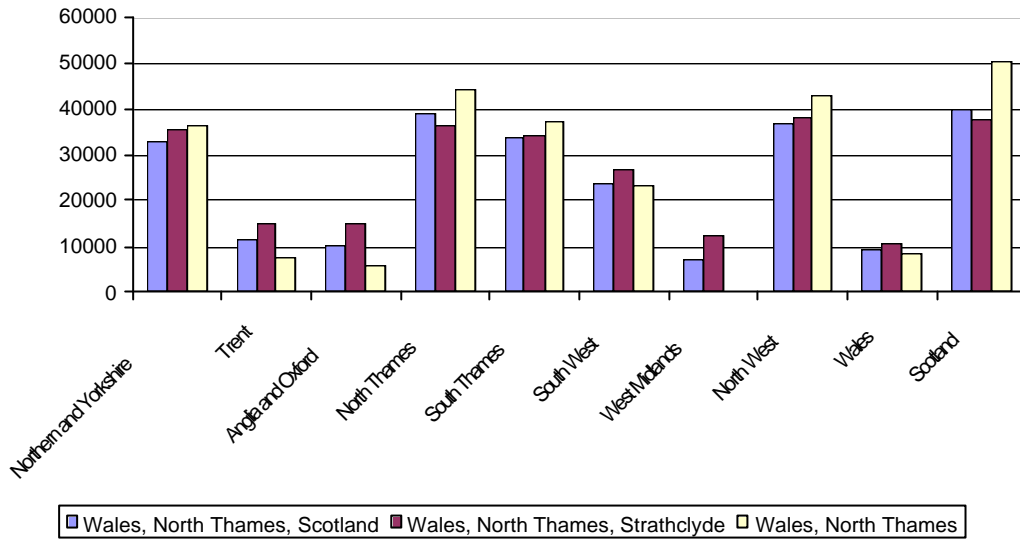


Figure 12: Effect of using only a part of a region as anchor point on regional estimates: anchor points Wales, North Thames, and Scotland or Strathclyde

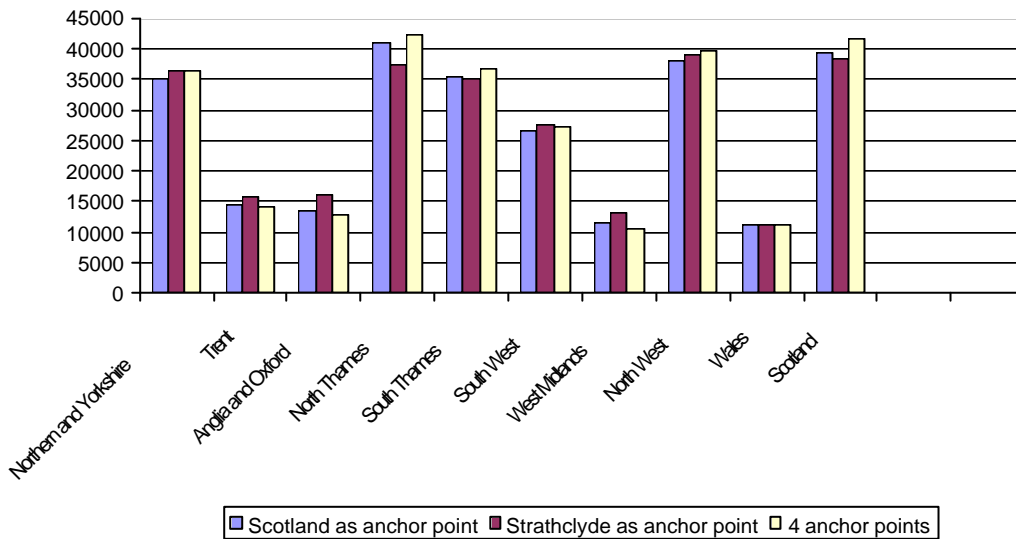


Figure 13: Effect of using only a part of a region as anchor point on regional estimates: anchor points Wales, North Thames, South Thames, West Midlands, and Scotland or Strathclyde

6.2 Violations of the requirements of the indicators

6.2.1 Selection of indicators

Often not all of indicators used in the data sets presented in chapter 4 are available. In this section we first analyse how the lack of some drug-related indicators influences the results. Then we examine if drug-related indicators can be replaced by social indicators.

Not all indicators available

The effect of the lack of one indicator was analysed with the western German and the UK data set. The national prevalence estimates are compared in table 11.

Table 11: Effects of choice of indicators on the national prevalence estimates

Data set	Prevalence estimates					
	All indicators	Without A	Without B	Without C	Without D	Without E
UK	265944	266033	274290	264126	275706	254579
Western Germany	97833	86355	81832	87791	118101	98488

The western German estimates range from 87,791 to 118,101, i.e. the deviation from the reference value of 97,833 ranges from -10% to 21% of the reference value. Compared to western Germany, the variation of the UK national estimates is rather small: the deviations of the reference value lie within an interval of -4% to 4% of the reference value.

The tables 12 and 13 present the regional results for Western Germany. Table 12 shows the results of the prevalence rates for all regions, table 13 the results of the ranked prevalence rates.

Table 12: Effects of choice of indicators on the prevalence rates of Western Germany

Region	Prevalence rates for the regions					
	All indicators	Without A	Without B	Without C	Without D	Without E
Baden-Württemberg	263	234	197	235	317	266
Bayern	215	180	129	184	278	224
Berlin	341	344	344	342	341	343
Bremen	495	481	24	490	497	514
Hamburg	546	557	483	550	544	525
Hessen	258	227	556	234	307	273
Niedersachsen	230	193	186	202	297	217
Nordrhein-Westfalen	271	238	24	245	327	266
Rheinland-Pfalz	214	173	163	171	281	229
Saarland	219	179	216	187	285	234
Schleswig-Holstein	203	168	126	157	276	215

A: offences; B: drug-related deaths, C: clients in treatment; D: AIDS cases and E: number of convictions

Table 13: Effects of choice of indicators on ranked prevalence rates of Western Germany

Region	Rank of prevalence rates for the regions					
	All indicators	Without A	Without B	Without C	Without D	Without E
Baden-Württemberg	5	5	5	5	5	5
Bayern	9	8	8	9	10	9
Berlin	3	3	3	3	3	3
Bremen	2	2	10	2	2	2
Hamburg	1	1	2	1	1	1
Hessen	6	6	1	6	6	4
Niedersachsen	7	7	6	7	7	10
Nordrhein-Westfalen	4	4	11	4	4	6
Rheinland-Pfalz	10	10	7	10	9	8
Saarland	8	9	4	8	6	7
Schleswig-Holstein	11	11	9	11	11	11

A: offences; B: drug-related deaths, C: clients in treatment; D: AIDS cases and E: convictions

As can be seen in both tables, the prevalence rates for some regions vary between the different models (e.g. Saarland), for some regions they don't vary (e.g. Baden-Württemberg). The greatest impact has the indicator B, representing the drug-related deaths. Table 14 picks out the region Hessen. Hessen ranks first for the calculation without indicator B (without drug-related deaths), and ranks fourths for the calculation without indicator E (without number of convictions) and is on the sixth place for all other calculations.

Table 14: Changes of prevalence rates for the region Hessen

Indicators	Hessen	
	Prevalence rate	Rank
Offences, drug-deaths, clients in treatment, AIDS cases, drug-related convictions	258	6
Without offences	227	6
Without drug-related deaths	556	1
Without number of clients in treatment	234	6
Without number of AIDS cases	307	6
Without number of convictions	273	4

Another example is Nordrhein-Westfalen (Table 15), which ranks 11th without drug-related deaths and on 6th place without number of convictions, whereas it is on the 4th place for all other calculations.

Table 15: Changes of Prevalence Rates for the Region Nordrhein-Westfalen

Indicators	Nordrhein-Westfalen	
	Prevalence rate	Rank
Offences, drug-deaths, clients in treatment, AIDS cases, drug-related convictions	271	4
Without offences	238	4
Without drug-related deaths	24	11
Without number of clients in treatment	245	4
Without number of AIDS cases	237	4
Without number of convictions	266	6

Tables 16 to 19 present the same calculations for the British data. It can be seen, that the choice of indicators has an impact on the regional prevalence rates and the order of the regions. Again, the indicator “drug-related deaths” is of great importance for the results, hinting at the significance of the content of the indicators.

Table 16: Effects of choice of indicators on the prevalence rates of the UK

Region	Prevalence rates for the regions					
	All indicators	Without A	Without B	Without C	Without D	Without E
Northern and Yorkshire	532	531	541	480	585	489
Trent	316	321	391	347	304	341
Anglia & Oxford	297	306	446	265	250	322
North Thames	568	562	474	640	615	539
South Thames	540	535	489	579	606	485
South West	435	435	440	479	475	405
West Midlands	224	230	346	222	191	301
North West	605	600	598	592	699	471
Wales	390	388	376	491	450	258
Scotland	768	774	805	557	627	807

A: convictions; B: seizure, C: treatment; D: AIDS and E: drug-related deaths

Table 17: Effects of choice of indicators on the ranked prevalence rates of the UK

Region	Ranked prevalence rates for the regions					
	All indicators	Without A	Without B	Without C	Without D	Without E
Northern and Yorkshire	5	5	3	6	5	3
Trent	8	8	8	8	8	8
Anglia & Oxford	9	9	6	9	9	9
North Thames	3	3	5	1	3	2
South Thames	4	4	4	3	4	4
South West	6	6	7	7	7	6
West Midlands	10	10	10	10	10	10
North West	2	2	2	2	1	5
Wales	7	7	9	5	6	7
Scotland	1	1	1	4	2	1

A: convictions; B: seizure, C: treatment; D: AIDS and E: drug-related deaths

To illustrate the results, the regions North Thames (Table 18) and North West (Table 19) have been selected.

Table 18: Changes of prevalence rates for the region North Thames

Indicators	North Thames	
	Prevalence rate	Rank
Drug-related convictions; seizure; number of clients; number of AIDS cases, drug-related deaths	568	3
Without number of convictions	562	3
Without seizure of controlled illegal drugs	474	5
Without number of clients in treatment	640	1
Without number of AIDS cases	615	3
Without drug-related deaths	539	2

Table 19: Changes of prevalence rates for the region North West

Indicators	North West	
	Prevalence rate	Rank
Drug-related convictions; seizure; number of clients; number of AIDS cases, drug-related deaths	605	2
Without number of convictions	600	2
Without seizure of controlled illegal drugs	598	2
Without number of clients in treatment	592	2
Without number of AIDS cases	699	1
Without drug-related deaths	471	5

It can be seen that the choice of a different set of indicators has an impact on the regional prevalence estimates. For both regions, different sets result in different variations. For the region North Thames the indicators “seizure of controlled drugs” and “clients in treatment” result in the greatest variations, for the region North West the indicator “drug-related deaths”.

Generally, the choice of the set of indicators seems to effect more the regional prevalence rates than the national one. For example, the number of drug-related deaths might differ to a great extent between different areas. Omitting this indicator results in a much lower rank for a region with a high death-rate in comparison to a model which includes this specific indicator. The fact that the national prevalence estimates do not differ immensely, underlines the above mentioned high influence of the selected anchor points.

Social indicators

Due to a lack of available drug-related indicators the Dutch work group (EMCDDA, 2000a) used an alternative model with social indicators. The objective of the following study was to analyse the effect of a set of social indicators. One of the original assumptions of the method is, that the indicators have to be related to drug use, a relationship which is rather unclear in the case of social indicators. A two-fold approach was chosen: firstly the Dutch model was applied to

German data, and secondly the theoretical implications were analysed. Table 20 shows the social indicators utilised for the multivariate indicator method for the Netherlands.

Table 20: Social Indicators and Anchor Points of the Multivariate Indicator Method for the Netherlands

Region	Population 15-54 years	A	B	C	D
Groningen	333080	100.54	496.89	4.97	860
Leeuwarden	348012	73.55	280.18	5.63	
Assen	257850	67.62	310.69	7.72	
Zwolle	442686	80.11	447.07	4.00	
Deventer	262279	116.92	557.85	3.11	
Almelo	341444	154.68	450.62	2.93	
Arnhem	523679	143.53	499.17	4.10	
Nijmegen	294148	179.37	523.83	4.34	
Amersfoort	287767	291.68	601.55	3.92	
Utrecht	347865	322.36	947.60	2.60	950
Amsterdam	457800	2173.90	1282.21	0.21	5800
Zaanstad	683587	429.97	568.57	2.70	
Hoorn	110864	196.55	462.58	6.41	
Den Helder	96490	87.46	404.70	5.19	
Alkmaar	137550	287.31	481.93	3.93	
Leiden	451597	482.24	467.84	3.32	
Den Haag	254877	2973.79	961.90	0.38	3300
Gouda	292310	318.71	473.46	4.13	
Dordrecht	401920	215.12	448.91	4.06	
Rotterdam	549981	1249.50	768.00	0.94	4000
Middelburg	203657	87.39	410.45	6.80	
Breda	616168	174.26	585.18	4.64	
Den Bosch	756446	177.85	562.81	4.08	
Venlo	270986	124.50	508.25	5.10	
Heerlen	377566	385.04	590.54	3.02	500

A =Housing density; B =Crimes against property; C =Mobility; D= Estimated size of regional IDU population (anchor points)

Social indicators were collected for Germany and the multivariate indicator method applied. The social indicators available were:

- Housing density (number of houses per square km)
- Crimes against property (number of crimes against property per 10,000 inhabitants)
- Mobility (migration within the municipality)
- Unemployed persons

Table 21 depicts the results for western Germany using a different subset of indicators. The first row shows the results for the drug-related indicators as a reference. The following rows present the results utilising the drug-related indicators plus one of the social indicators. A model combining all drug-related and social indicators has been calculated, as well as a model with the original three social indicators utilised by the Netherlands and a model with four social indicators.

Table 21: Changes in Prevalence Rates with different indicators including social indicators for western Germany

Indicators used	Variables	Western Germany	
		Prevalence rate	Difference
Offences, drug deaths, number of clients, number of AIDS cases, drug-related convictions = drug-related indicators	A-E	97,833	Reference
Drug-related indicators + unemployed persons	A-E + K	78,345	-19,488
Drug-related indicators + housing density	A-E + H	103,145	5,312
Drug-related indicators + crimes against property	A-E + I	81,318	-16,515
Drug-related indicators + mobility	A-E + J	72,514	-25,319
Drug-related indicators + housing density + crimes against property + mobility + unemployed persons	A-E + H-K	61,059	-36,774
Housing density, crimes against property, mobility, unemployed persons = four social indicators	H-K	210,470	112,637
Housing density, crimes against property, mobility = three social indicators	H-J	147,853	50,020

It can be seen from table 21 that there are dramatic changes depending on the subset of indicators used. The greatest differences can be observed for the use of social indicators without drug-related indicators. The application with the original three social indicators raises the prevalence rate by 50%, whereas the application with four social indicators doubles the prevalence rate. This is even more striking when considering the results of the sensitivity analysis above, as it was shown, that the anchor points have a much greater impact on the national prevalence estimate than the choice of indicators. All calculations of Table 21 have been conducted with the same anchor points.

Table 22 presents the results of the above calculations for the prevalence rates of the regions ordered by size.

Table 22: Prevalence rates of the regions, ordered by size

Region	Rank of prevalence rates for the regions					
	A-E	A-K	A-E + unem- ployment	A-E + housing	A-E + crimes	A-E + mobility
Baden-Württemberg	5	8	6	5	6	6
Bayern	9	11	11	9	11	9
Berlin	3	3	3	3	3	3
Bremen	2	1	2	1	2	2
Hamburg	1	2	1	2	1	1
Hessen	6	5	5	6	5	5
Niedersachsen	7	7	7	10	7	8
Nordrhein-Westfalen	4	4	4	4	4	4
Rheinland-Pfalz	10	10	9	8	10	10
Saarland	8	6	8	7	9	7
Schleswig-Holstein	11	9	10	11	8	11

A-E: drug-related indicators; H-K: social indicators

Table 22 shows, that there is indeed a difference in the rank order of the prevalence rates of the different regions, as could be concluded from the sensitivity analysis. However, only the results including drug-related indicators are presented. Table 23 presents the model applied to the social indicators. Two models are compared: the second subset of indicators additionally includes “unemployment”.

Table 23: Effects of a different subset of social indicators for Western Germany

Regions	Social indicators (housing density, crimes against property, mobility)	Social indicators (plus unemployment)
	Prevalence rate	Prevalence rate
Baden-Württemberg	376	593
Bayern	379	590
Berlin	450	464
Bremen	480	429
Hamburg	451	488
Hessen	390	571
Niedersachsen	386	557
Nordrhein-Westfalen	401	545
Rheinland-Pfalz	379	582
Saarland	400	543
Schleswig-Holstein	385	571

Table 24 shows great differences in the prevalence rates for the regions dependent on the subset of social indicators used. This becomes even more clear, when the ranks of the regions are presented.

Table 24: Changes of prevalence rates for the region North West

Länder	Social indicators (housing density, crimes against property, mobility; unemployment)	Social indicators (housing density, crimes against property, mobility)
	Rank of prevalence rate	Rank of prevalence rate
Baden-Württemberg	1	11
Bayern	2	10
Berlin	10	3
Bremen	11	1
Hamburg	9	2
Hessen	5	5
Niedersachsen	6	7
Nordrhein-Westfalen	7	4
Rheinland-Pfalz	3	9
Saarland	8	5
Schleswig-Holstein	4	8

Table 24 shows, that the rank order of the regions turns almost completely to a reverse order with the use of “unemployment” as additional social indicator. Baden-Württemberg ranks first for all four social indicators and last for only three of them, Bayern ranks second with unemployment and 10th without. The results are even more striking when considering that both models are calculated with the same set of anchor points. The empirical results are not in favour of using social indicators, at least not for Germany. Furthermore, what are the theoretical assumptions underlying the use of a set of indicators?

When applying the multivariate indicator method to social indicators, such as housing density, mobility, crimes against property, and unemployment, the following questions have to be considered:

- Which target group is estimated by social indicators? What relationship is there between social indicators and drug prevalence? Is there enough empirical basis to conclude that these factors have a monotonous relationship with drug use?
- What about cannabis and alcohol? How can they be excluded? If there is a monotonous relationship to drug use, which are the drugs used by this population? How can a certain target group be specified? The definition agreed upon for national estimates explicitly excludes mere consumption of cannabis and alcohol. Can this be done for social indicators as well?
- Face validity: What is the evident relationship between unemployment and drug prevalence?

The interpretation of the common, underlying factor of mobility and housing density as drug prevalence is not evident. Another possible interpretation would be “aggression”. It is striking that the additional factor “unemployment” results in such great differences. What are the empirical concepts connected to unemployment? The following psychological constructs are correlated to unemployment:

- Educational level
- Anxiety
- Depression
- Grief
- Suicidality
- Family problems
- Divorce
- Reduced self-esteem
- Aggression
- Dependence on psychotropic substances
- Not substance-induced dependence (gambling, eating)
- Reduced coping-abilities
- Reduced general condition
- More health problems

It can be seen that drug use is correlated to unemployment, however, at least for Germany, a multitude of other factors are connected, too. It is not evident that unemployment contributes much to drug use. In general, the use of social indicators needs a theoretical foundation.

6.2.2 Systematic biases

Delays in data entry, possible inclusion of previous drug users, e.g. in the case of AIDS cases related to i.v. drug use, or report of event-based data instead of person-based data are probable systematic biases of indicators. In this section the impact of systematic biases on the prevalence estimates is analysed. In the following, the analysis is described for the case of systematic biases due to the use of event-based data sources in some regions. Note, that the results are applicable to all situations where some or all regional indicator values are biased. Note further, that a uniform bias in the sense that the observed indicator values are a multiple of the true values does not affect the prevalence estimates as the standardized values do not change.

Person-based data sources include individual data, event-based data sources count events, thus and include data of the same person more than once. This abstract concept is best illustrated by police data, which may count the number of drug-related convictions, but not the number of convicted persons. An individual could be counted twice or more often if he is convicted more than once in the referred period. What happens if in some administrative areas data are coded individually, in other administrative areas event-based? Should this indicator be omitted or is the calculation still valid? This effect was analysed with the German and the British data. The mixture

of event-based and person-based data sources was simulated by multiplying one of the indicators with five for around half of the regions. Different models were calculated, repeating the multiplication for all indicators. The results were then compared on a regional level and for the total prevalence. Table 25 shows the prevalence rates for the regions and the whole of Germany.

Table 25: Prevalence rates for western Germany and national prevalence estimate

Region	All indicators	Prevalence rates for the regions				
		5* A	5* B	5* C	5* D	5* E
Multiplied with five for the first five regions						
Baden-Württemberg	263	257	256	256	269	261
Bayern	215	208	205	207	221	213
Berlin	341	341	341	341	341	341
Bremen	495	493	494	494	495	497
Hamburg	546	547	547	547	546	545
Hessen	258	231	228	231	246	246
Niedersachsen	230	205	212	206	232	202
Nordrhein-Westfalen	271	240	248	240	268	243
Rheinland-Pfalz	214	191	190	184	215	208
Saarland	219	195	189	195	219	211
Schleswig-Holstein	203	186	176	173	208	195
National prevalence	97833	91082	91590	90649	98086	92981

A: offences; B: drug-related deaths, C: clients in treatment; D: AIDS cases and E: number of convictions

It can be seen that prevalence rates slightly change. Contrary to the above findings, the effect of partially event-, partially person-based indicators is not dependent on the content of the indicator. There is no great effect for the indicator “drug-related deaths” in comparison to the other indicators. In consequence, the national prevalence estimates vary within a small range of –7.5% to 0.3% of the reference value 97,833. Table 26 shows the regions ordered by prevalence rate.

Table 26: Western German regions ordered by prevalence rate

Region	All indicators	Rank of prevalence rates for the regions				
		5* A	5* B	5* C	5* D	5* E
		Multiplied with five for the first five regions				
Baden-Württemberg	5	4	4	4	4	4
Bayern	9	7	8	7	8	8
Berlin	3	3	3	3	3	3
Bremen	2	2	2	2	2	2
Hamburg	1	1	1	1	1	1
Hessen	6	6	6	6	6	5
Niedersachsen	7	8	7	8	7	10
Nordrhein-Westfalen	4	5	5	5	5	6
Rheinland-Pfalz	10	10	9	10	10	9
Saarland	8	9	10	9	9	7
Schleswig-Holstein	11	11	11	11	11	11

If the regions are ordered by prevalence rate, it can be seen that the effect is a very slight one, as the order does not change dramatically. The same calculations were conducted with the British data. Table 27 shows the prevalence rates for the UK and its regions, table 28 shows the regions ordered by prevalence rate. In the British data, the effect on the regional prevalence rates seems to be greater than in the German data. The national prevalence estimates, however, lie within an interval of -4.4% to 3.0% of the reference value 265,944.

Table 27: Prevalence rates for the UK regions and national prevalence estimate

Region	All indicators	Prevalence rates for the regions				
		5* A	5* B	5* C	5* D	5* E
		Multiplied with five for the first five regions				
Northern & Yorkshire	532	539	532	541	533	478
Trent	316	324	398	398	324	341
Anglia & Oxford	297	300	457	331	278	325
North Thames	568	581	462	640	665	523
South Thames	540	550	479	601	585	472
South West	435	426	447	469	436	414
West Midlands	224	218	365	274	200	319
North West	605	590	591	560	608	476
Wales	390	380	385	484	415	367
Scotland	768	759	799	491	625	808
National prevalence	265944	266069	273970	271316	267964	254164

Table 28: Person- vs. event-based data sources; prevalence rates ordered by size

Region	All indicators	Rank of prevalence rates for the regions				
		5* A	5* B	5* C	5* D	5* E
Multiplied with five for the first five regions						
Northern & Yorkshire	5	5	3	4	5	3
Trent	8	8	8	8	8	8
Anglia & Oxford	9	9	6	9	9	9
North Thames	3	3	5	1	1	2
South Thames	4	4	4	2	4	5
South West	6	6	7	7	6	6
West Midlands	10	10	10	10	10	10
North West	2	2	2	3	3	4
Wales	7	7	9	6	7	7
Scotland	1	1	1	5	2	1

Although a slight effect of a changed order of regions can be observed, the method seems to be relatively robust towards this data problem.

6.2.3 Indicator not reported by area of residence

Usually police statistics of arrests for possession of illicit drugs are not reported by area of residence but by area of report. If a metropolis is situated near the border to another region, residents of the neighbouring region have a high probability of being arrested in the city because drug trafficking is more concentrated in this area. The impact of this problem is analysed with the German data set where all three anchor points – Berlin, Bremen and Hamburg - are metropolis by lessening the offences in the metropolis and increasing the number in adjacent regions. As Berlin is surrounded by eastern German regions, we did not change the indicator value of Berlin. Bremen is enclosed by Niedersachsen, and Hamburg is adjoined both by Niedersachsen and Schleswig-Holstein. Thus, we shifted 10% of the convictions in Bremen from Bremen to Niedersachsen. The number of convictions in Hamburg was also lessened by 10%. These cases were distributed equally to Niedersachsen and Schleswig-Holstein. The figure of 10% was chosen arbitrarily. Table 29 summarizes both the original and the new indicator values, the indicator rates per 100,000 15-54-year old inhabitants, and the prevalence estimates. It can be seen, that shifting of cases leads to negligible changes both of regional and national prevalence estimates. The national estimate changes from 97,834 to 97,269.

Table 29: Effect of shifting cases between regions: western Germany, 1995

Western German regions, 1995	Population 15-54 years	Law enforcement		Law enforcement (population standardized)		Prevalence estimates	
		Original	New	Original	New	Original	New
Baden-Württemberg	5774803	13225	13225	229.01	229.01	15191	15103
Bayern	6694822	9538	9538	142.47	142.47	14373	14119
Berlin	2047829	5507	5507	268.92	268.92	6982	6983
Bremen	376691	2689	2419	713.85	642.17	1866	1865
Hamburg	973412	6827	6147	701.35	631.49	5312	5314
Hessen	3384030	7825	7825	231.23	231.23	8733	8682
Niedersachsen	4262229	8020	8630	188.16	202.48	9794	9775
Nordrhein-Westfalen	9812031	26759	26759	272.72	272.72	26629	26581
Rheinland-Pfalz	2161042	3594	3594	166.31	166.31	4621	4550
Saarland	588738	1043	1043	177.16	177.16	1289	1271
Schleswig-Holstein	1501615	1585	1925	105.55	128.20	3044	3026
Total	37577242	86612	86612			978343	97269

6.2.4 Influence of the range of the age group

As indicators are often not broken down by age group the choice of the age range is rather arbitrary. To examine the influence of the range of the age group on national and regional prevalence estimates we standardized the German data both to the size of the 15-39-year-old population and to the 15-54-year-old-population. As can be seen from figure 14, there are hardly any differences between the corresponding regional estimates. The national estimates are 99,659 (15-39-year-olds) and 97,833 (15-54-year-olds). This small effect was expected as in each region the population size of the 15-54-year-olds is about 1.5 times the population size of the 15-39-year-olds. Thus, standardizing (subtracting the mean and dividing by the standard deviation) the population related indicators yields almost the same figures. Unless there are major difference between the regional age distributions the selection of the age range has no substantial impact on the prevalence estimates. As a consequence, to get prevalence estimates of a certain age group a breakdown of the indicators and the anchor point estimates by age group is needed.

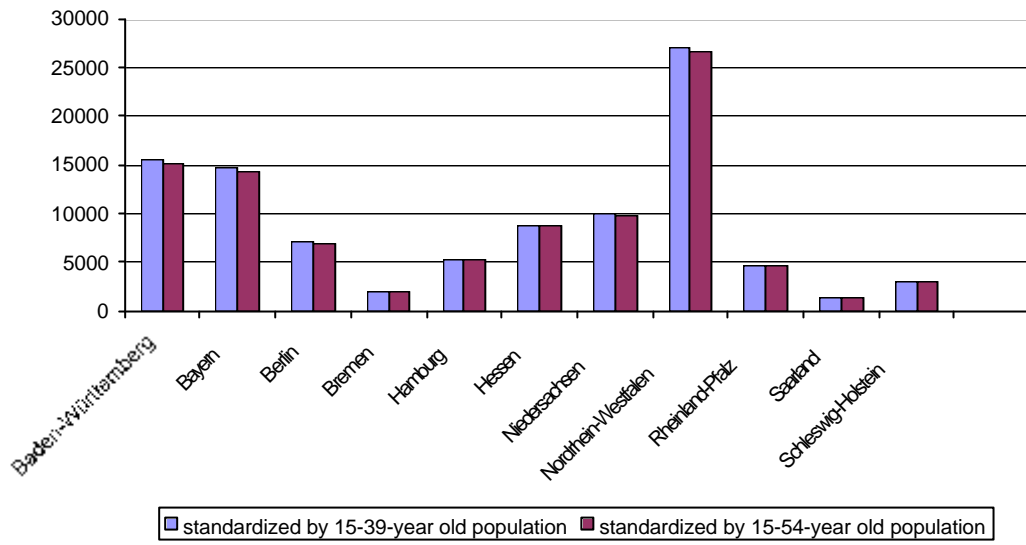


Figure 14: Influence of the age group on regional prevalence estimates

7 Comparison of variants of the multivariate indicator method

The multivariate indicator method can be seen as an extension of the synthetic estimation by linear extrapolation of a proxy variable. Given reliable prevalence estimates and the values of a proxy variable for two cities (anchor points) as well as the value of the proxy variable for a third city, the drug use prevalence of the third city can be estimated by linear regression with the proxy variable as independent variable and the prevalence estimates of the anchor points as dependent variable (Wickens, 1993). The generalisation to national prevalence estimation is straightforward: The prevalence of all communities or regions has to be estimated. The sum of all regional prevalence estimates yields the national prevalence estimate.

Obviously, the validity of prevalence estimation can be improved by increasing the number of anchor points. Then, more drug use indicators (proxy variables) can be used in the linear regression model. One of the problems is, however, the choice of appropriate drug use indicators (proxy variables). If the number of drug use indicators equals or exceeds the number of anchor points linear regression is not possible. As drug use indicators are more easily available than reliable regional prevalence estimates it is often necessary to reduce the number of drug use indicators. Up to now, different methods of reducing the number of indicators have emerged:

- Mariani (1999) as well as Person, Retka and Woodward (1977, 1978) applied a principal component analysis (PCA). While Mariani calculated the PCA with the drug use indicators per 100,000 inhabitants Person, Retka and Woodward used the ranks of the drug use indicators.
- Brugal et al. (1999) summed the ranks of the various drug use indicators in neighbourhoods in Barcelona. This was also proposed by Person, Retka and Woodward (1977).
- During a meeting of the project “Study to obtain Comparable National Estimates of Problem Drug Use Prevalence for all EU Member States” Rossi (1998) proposed to base the choice of the indicators on the correlation matrix of the indicators. In the case of three anchor points the two indicators with the smallest correlation should be used.
- In their analysis of construct validity of the multivariate indicator and the capture-recapture method Woodward et al. (1984) used both the original indicator values and the natural logs of the per capita occurrence of the indicators (indicators divided by population size). The aim of the log-transformation was to normalise the platykurtic distribution observed for the rates.

Summing up, the multivariate indicator method has different variants which do not only differ in the method of reducing the number of indicators (principal component analysis, based on correlation matrix, summing up) but also in the method of transforming the indicators (ranking the original indicators, calculating the rate per 100,000 inhabitants, taking the logs of the rate per inhabitant or using the original indicators, i.e. no transformation).

7.1 Different variants applied to actual data sets

To get a better idea of the properties of different variants we analysed the UK data set, the 1995 Italy data set, as well as the western German data set with the following independent variables in the regression analysis step:

- First principal component of the original indicators (PCA of original data)
- First principal component of the indicator rates (PCA per 100,000)
- First principal component of the ranked indicators (PCA of ranks)
- First principal component of the ranked indicator rates (PCA of ranks per 100,000)
- Original indicators with the lowest correlation (Correlation of original data)
- Indicator rates with the lowest correlation (Correlation per 100,000)

Only linear regression was applied. Note, that dependent and independent variables must fit to each other. If the independent variables are rates, the prevalence estimates of the anchor points have to be transformed to prevalence rates as well. Negative prevalence estimates were replaced by 0.

Tables 30-32 give the prevalence estimates obtained by the six variants of the multivariate indicator method. In the Italian data set the estimates ranged between 301,841 (correlation per 100,000) and 476,934 (PCA of original data). The estimate obtained by correlation of original data was 344,874. The other variants yielded rather similar estimates, ranging between 372,162 and 386,112. In the western German data set all three anchor points were regions with a rather high prevalence rate per 100,000 inhabitants. Here the estimates varied between 24,938 (PCA with ranks per 100,000) and 741,574 (correlation per 100,000)! The other results were 61,926 (PCA with ranks per 100,000), 63,391 (correlation of original data), 72,845 (PCA of original data), and 97,834 (PCA per 100,000). Compared to the results for Italy and western Germany, the variation of the UK national prevalence estimates was rather small, ranging from 231,038 to 278,529. Moreover, except for the variant “correlation per 100,000”, the resulting estimates are within a small range of about 260,000 to 280,000.

Table 30: Estimates obtained by different variants: Italy, 1995

Regions	PCA (original data)	PCA (per 100,000)	PCA with ranks (original data)	PCA with ranks (per 100,000)	Correla- tion (original data)	Correla- tion (per 100,000)
Abruzzo	17900	7984	6304	7371	11924	9984
Basilicata	14815	2892	0	2279	2134	1105
Calabria	17010	9491	4609	7508	4642	2036
Campania	31427	36148	36798	31805	37077	27309
E. Romagna	30271	28970	35748	30148	23669	23022
Friuli	17054	6871	3203	6349	6285	4758
Lazio	38234	42426	44526	42558	53733	53733
Liguria	21352	12893	19973	12892	9297	8210
Lombardia	51302	67393	50050	67399	42297	42297
Marche	17302	7826	6001	7244	5189	4017
Molise	14279	1481	0	1093	1953	1398
Piemonte	33922	34611	42012	35852	40731	38126
Puglia	25806	25307	29035	24030	25637	21582
Sardegna	20112	12293	15072	12246	13618	13618
Sicilia	23327	24435	23473	18880	18903	13972
Toscana	27172	24431	32540	24956	18885	16611
Trentino	17011	6371	3543	6268	3340	1886
Umbria	16400	5090	0	5172	5605	4678
Valle d'Aosta	14096	805	0	814	1541	1507
Veneto	28142	28394	34623	27298	18414	11993
Total	476934	386112	387510	372162	344874	301842

Table 31: Estimates obtained by different variants: western Germany, 1995

Regions	PCA (original data)	PCA (per 100,000)	PCA with ranks (original data)	PCA with ranks (per 100,000)	Correla- tion (original data)	Correla- tion (per 100,000)
Baden-Württemberg	11727	15191	10091	936	9228	113463
Bayern	8951	14373	8228	0	6101	142947
Berlin	6807	6982	6190	6990	7073	7073
Bremen	1730	1866	1821	1912	1798	1798
Hamburg	5778	5312	6304	5190	5444	5444
Hessen	6698	8733	6686	2361	4488	58207
Niedersachsen	6016	9794	6929	0	7365	104737
Nordrhein-Westfalen	20876	26629	11267	7549	18186	208257
Rheinland-Pfalz	2489	4621	3024	0	1583	50011
Saarland	437	1289	0	0	916	12978
Schleswig-Holstein	1336	3044	1386	0	1209	36659
Total	72845	97834	61926	25038	63391	741574

Table 32: Estimates obtained by different variants: UK

Regions	PCA (original data)	PCA (per 100,000)	PCA with ranks (original data)	PCA with ranks (per 100,000)	Correla- tion (original data)	Correla- tion (per 100,000)
Northern and Yorkshire	37395	35095	35856	33172	39331	33867
Trent	15452	14574	18229	11675	13629	18127
Anglia and Oxford	12391	13426	15901	12278	10449	0
North Thames	46979	40825	45823	42980	44410	44410
South Thames	39424	35510	40678	42489	38140	38140
South West	29923	26676	27886	28591	27858	37007
West Midlands	14915	11524	14902	10943	13130	13130
North West	38849	37944	38621	35252	34637	0
Wales	9296	11064	8924	10032	8357	8357
Scotland	31423	39307	31710	34784	38000	38000
Total	276047	266045	278529	262195	267941	231038

As can be seen immediately, the variation of the prevalence estimates within most of the regions exceeds the variation of the national prevalence estimates. With regard to Italy, e.g., the highest national prevalence estimate is 1.6 times the lowest national prevalence estimate, while this factor rises to 9.0 in Trentino or to 8.4 in Calabria. PCA of the ranks yielded negative prevalence estimates for four of the twenty Italian regions, which were replaced by zero. As this methods result in rather high prevalence estimate in other regions the corresponding national prevalence estimate compares with the national prevalence estimates obtained by PCA per 100,000, PCA with ranks per 100,000, and correlation of the original data.

The within region variation of the western German regions is larger than observed in the Italian data. The high estimates obtained by the variant “correlation per 100,000” can be explained by the negative correlation of the two indicator rates involved in the estimation procedure (see figure 15). The problem of a negative correlation can be ruled out if the anchor points are not too similar in prevalence or in prevalence rate. If high prevalent regions and low prevalent regions anchor points are selected as anchor points negative correlations between indicator (rates) and prevalence (rate) estimates are unlikely. On the other hand, the indicators or indicator rates used in the correlation variants should be selected among the positively correlated indicators or indicator rates.

Contrary to the Italian and German results, the different variants applied to the UK data yield rather similar regional estimates. The only exceptions are Anglia and Oxford as well as North West where the variant “correlation per 100,000” resulted in a negative prevalence estimate, which was replaced by zero.

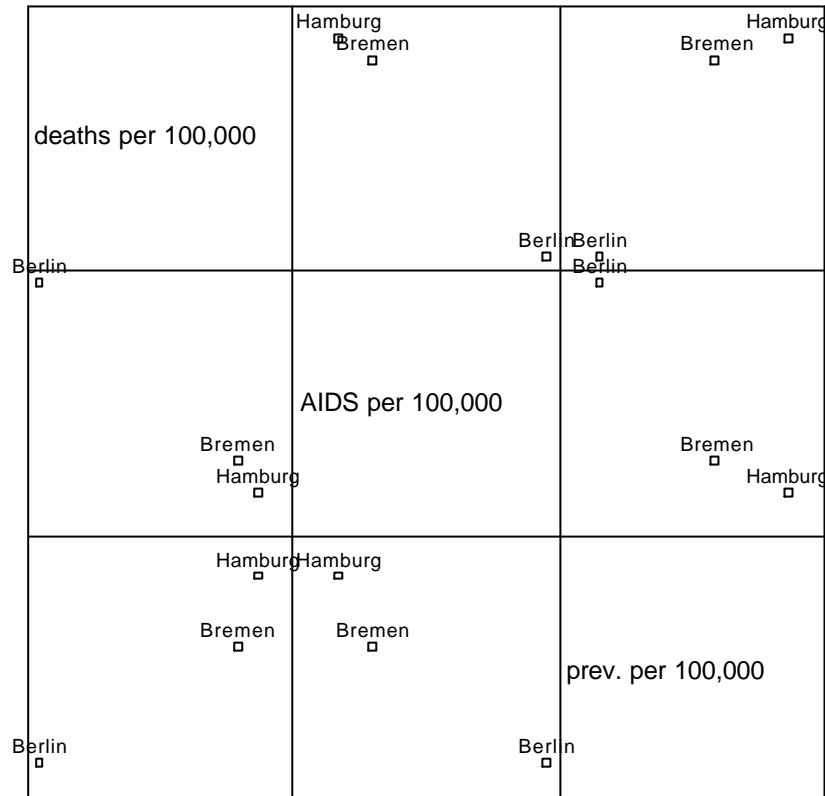


Figure 15: Scatterplot of indicator rates and prevalence rates used in correlation per 100,000 in western Germany

Moreover, the influence of the anchor points on the different variants was analysed by omitting anchor point estimates in the UK data set. We estimated the prevalence both based on the four anchor points with the highest prevalence estimates (i.e. without the West Midlands) and based on the four anchor points with the lowest prevalence estimates (i.e. without North Thames). Note, that the lowest and highest anchor point estimates per 100,000 inhabitants are found in Wales and in Scotland respectively. Thus, the selection of anchor points are the worst four-anchor-point cases for the variants based on the original data, but not for the variants based on the rates per 100,000 inhabitants.

The correlation variants were calculated with four independent variables in the case of five anchor points and with three independent variables in the case of four anchor points. The variant “correlation with original data” was based on the indicators convictions, seizures, clients in treatment, and AIDS related to IDU in the case of five anchor points. Seizures were omitted in the case of four anchor points. In the case of “correlation per 100,000” the indicators convictions, seizures, AIDS related to IDU, and drug-related deaths, and convictions, seizures, and drug-

related deaths were employed. In the case of three anchor points the indicators AIDS related to IDU and drug-related deaths were used. Tables 33 and 34 show the correlation matrices.

Table 33: Correlation matrix of the indicators used in Great Britain

	Convictions	Seizures	Clients in treatment	AIDS	Drug-related deaths
Convictions	1	.786	.527	-.141	.754
Seizures		1	.760	.461	.782
Clients in treatment			1	.421	.885
AIDS				1	.150
Drug-related deaths					1

Table 34: Correlation matrix of the indicators per 100,000 inhabitants in Great Britain

	Convictions	Seizures	Clients in treatment	AIDS	Drug-related deaths
Convictions	1	.453	-.136	-.494	.255
Seizures		1	.539	.501	.425
Clients in treatment			1	.599	.644
AIDS				1	.114
Drug-related deaths					1

Table 35: National prevalence estimates for Great Britain obtained by different variants of the multivariate indicator method, employing different anchor points

Anchor Points	PCA original data	PCA per 100,000	Correlation original data	Correlation per 100,000
North Thames, South Thames, West Midlands, Wales, Scotland	276047	265944	267941	231038
South Thames, West Midlands, Wales, Scotland	286339	259042	261292	276571
North Thames, South Thames, West Midland, Scotland	280604	280997	263536	278650
North Thames, West Midlands, Scotland	284124	277858	289728	362063
North Thames, South Thames, Scotland	354194	302704	268081	325063
South Thames, West Midlands, Wales	262566	262185	398986	408420

Omission of one anchor point has only a substantial influence on the national prevalence estimates obtained by the variant “correlation per 100,000” (table 35). Here the estimates rise from about 230,000 to about 280,000 if only four anchor points are employed. This variant performs also worst in the case of three anchor points, where the estimate increases further to 408,000 if South Thames, West Midlands and Wales are used as anchor points. Application of the variant “correlation of original data” with these anchor points leads also to a national prevalence estimate

of about 400,000, while the variant PCA per 100,000 yields an estimate very close to the result based on five anchor points. With regard to the PCA-variants, the highest deviations from the results based on five anchor points were found in the case of the anchor points North Thames, South Thames, and Scotland. In the case of correlation of the original data, the same combination of anchor points, however, results an estimate close to the estimate based on five anchor points.

The regional estimates obtained using four anchor points are compared in figure 16. It shows the regional prevalence estimates if the anchor point estimate for Wales, which is the anchor point with the lowest prevalence, is not used. Omission of this anchor point had a higher influence on the correlation-variants than on the PCA-variants. Whereas no regional estimate obtained by the PCA-variants deviates substantially from those in the five anchor point-situation, e.g. the regional estimate for Trent obtained by “correlation with original data” declines from 13629 to 9210. The estimates obtained by “correlation per 100,000” are more similar to the results of the PCA with original data than in the five anchor point-situation. Negative prevalence estimates did not emerge.

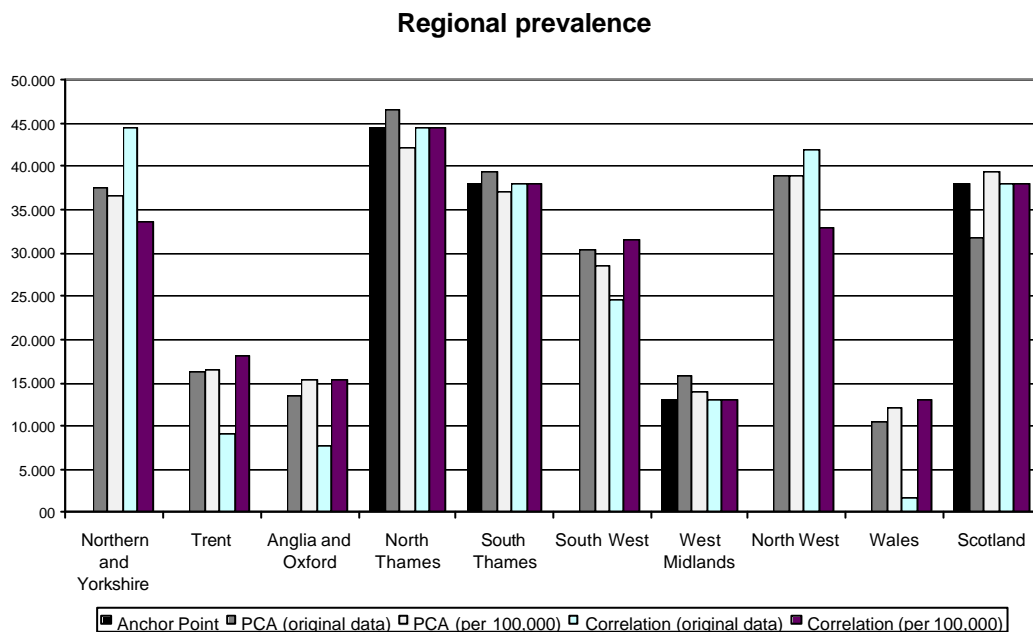


Figure 16: Regional prevalence estimates for Great Britain obtained by different variants of the multivariate indicator method, employing the anchor points North Thames, South Thames, West Midlands, and Scotland

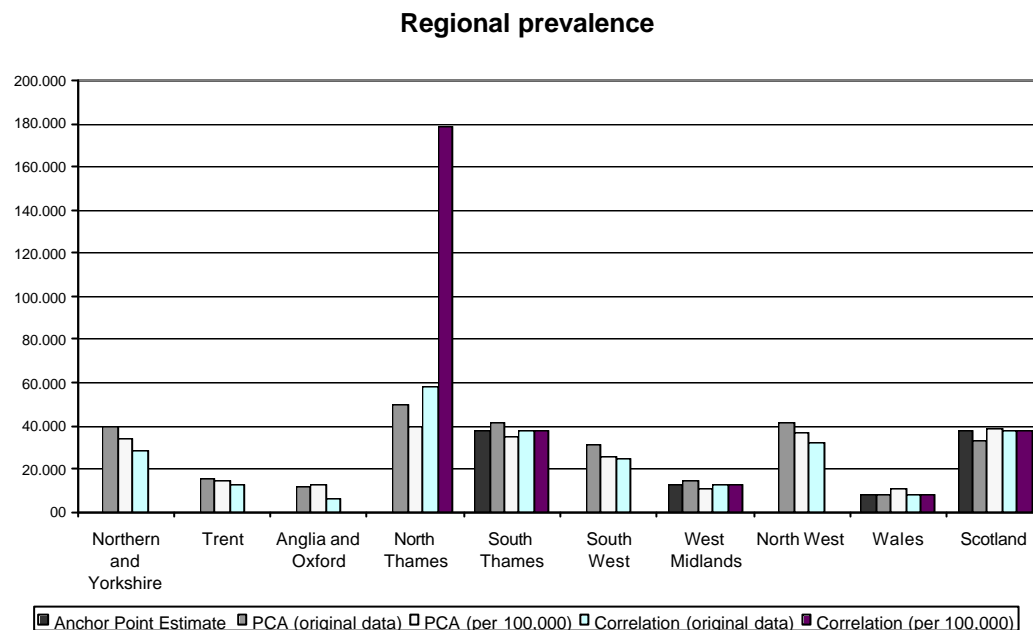


Figure 17: Regional prevalence estimates for Great Britain obtained by different variants of the multivariate indicator method, employing the anchor points South Thames, West Midlands, Wales, and Scotland

Figure 17 depicts the regional prevalence estimates if the anchor point estimate for North Thames is not employed. As can be seen immediately, omission of this anchor point increases the estimate obtained by the variant “correlation per 100,000” from 44,410 to about 180,000. All the other estimates for the non-anchor-point regions obtained by this variant were negative and were therefore replaced by zero. Again, the influence on the PCA-variants is small. The variant “correlation with original data” yields for some regions rather different estimates compared to the five anchor point-situation, e.g. the estimate for Northern and Yorkshire declines from 39,331 to 28,732.

7.2 Simulation

The empirical results do not indicate which variant is the best. It was decided to simulate different situations to analyse bias, variance and mean square error of the different variants of the multivariate indicator method. A variant which performs well in all or most of the simulated situations is regarded as an appropriate method for prevalence estimation.

Up to now, only one very simplistic situation has been simulated. In the simulated situation with 7 low prevalent regions, 7 medium prevalent regions and 6 high prevalent regions each region had 100,000 inhabitants. Thus, the indicator rates are the same as the indicator values and the number of variants of the multivariate indicator method reduces to three. The simulated indicators were

convicts, drug-related deaths, addicts in treatment, AIDS cases as well as addicts in jail. All these indicators were normal distributed. Within each type of region (low, medium, and high prevalent) the same mean was chosen for the different indicators. E.g, in all low prevalent regions the mean of the convicts was 100, while the mean of drug-related deaths was 4 (Table 36).

Table 36: Simulated situation

Variables	Region 1 - 7	Regions 8 – 14	Regions 15 - 20
Convicts	N (100; 100)	N (150; 100)	N (200; 100)
Deaths	N (4; 1)	N (5; 1)	N (6; 1)
Treatment	N (300; 400)	N (400; 400)	N (500; 400)
AIDS	N (6; 1)	N (7; 1)	N (8; 1)
Jail	N (40; 16)	N (50; 16)	N (60; 16)
Prevalence	3000	4000	5000

One thousand simulations were run. Bias, variance as well as the mean square error and the maximum prevalence estimate were analysed for each of the regions. As can be seen from figures 18 to 21, the variant “correlation” performs badly in all regions except for the anchor points. The variant “principal components of original indicator values” was superior to the variant “principal components of ranked indicator values” in the low prevalent regions and the medium prevalent regions, but performed worse in the high prevalent regions.

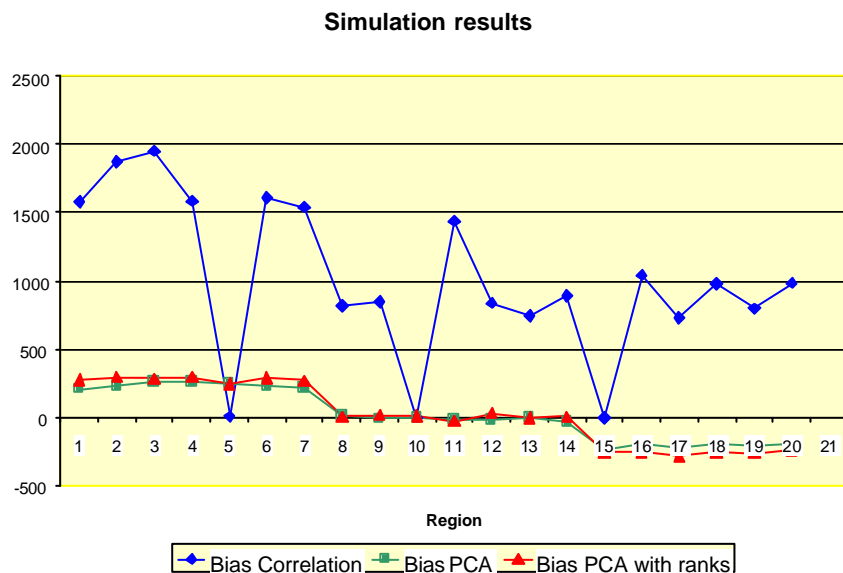


Figure 18: Bias of all three simulated variants

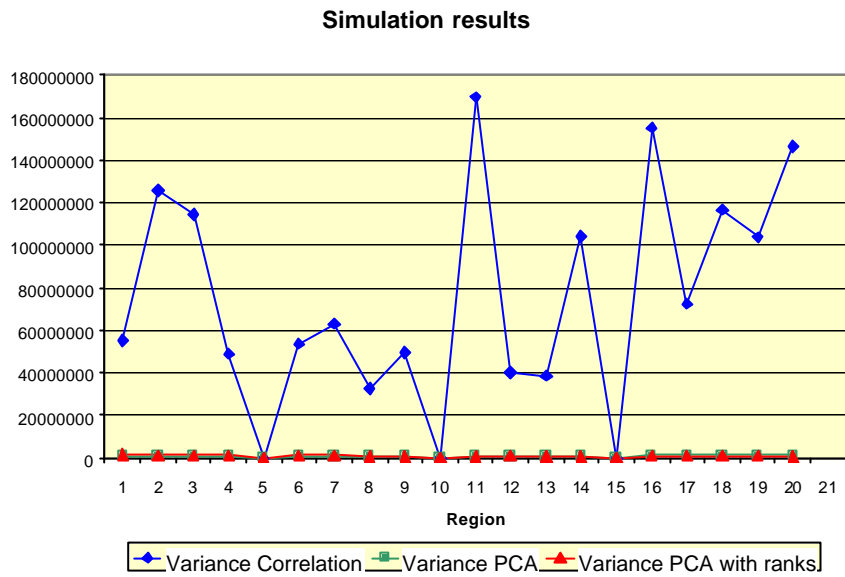


Figure 19: Variance of all three simulated variants

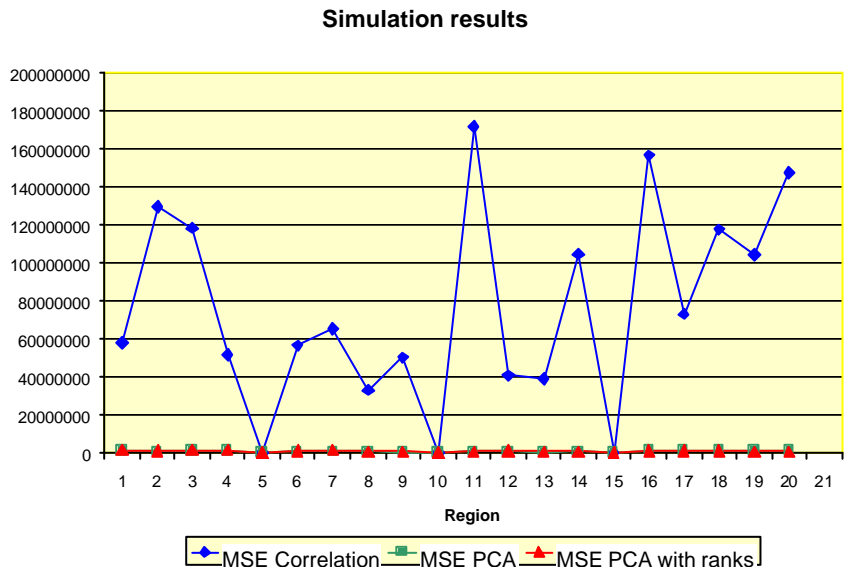


Figure 20: Mean square error of all three simulated variants

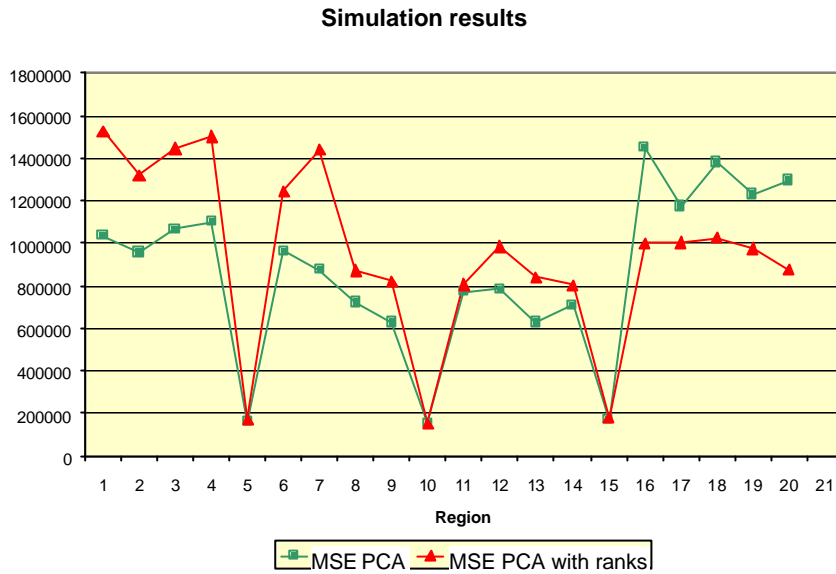


Figure 21: Mean square error of the variants based on principal component analysis

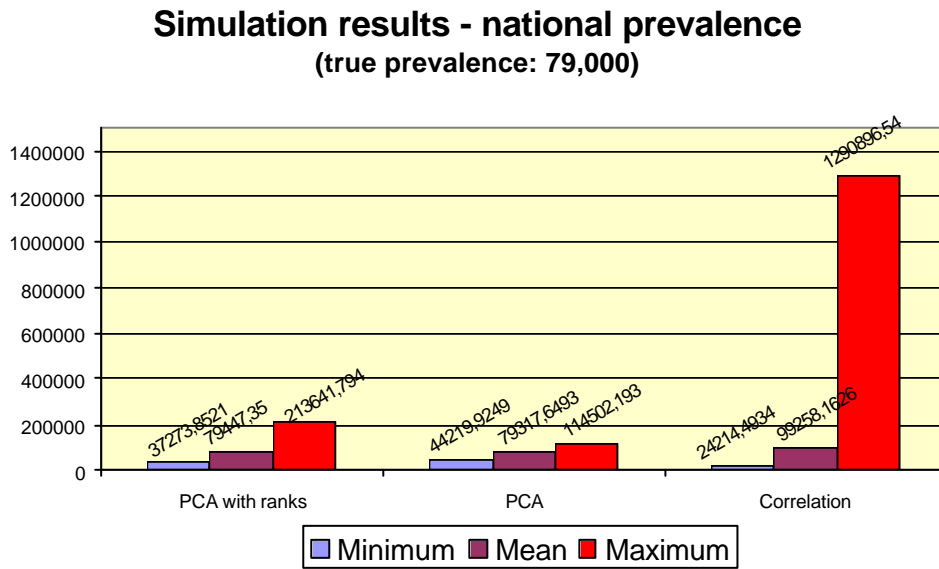


Figure 22: Minimum prevalence estimate, mean prevalence estimate, and maximum prevalence estimate obtained by the three variants

In the simulated situation principal component analysis based on the indicator values turned out to be the best method for national prevalence estimation: Both the range of the estimates and the deviation of the mean prevalence estimates from the true prevalence are smallest (Figure 22).

In further simulation studies more realistic assumptions should be made which should be derived from the existing data sets. This includes both the univariate distribution of indicator values (skewness, variance, mean, type of distribution) and the correlation between the indicators.

8 Cross-validation

The effect of different violations of the assumptions of the multivariate indicator method is analysed in chapter 6. It was, however, not possible to examine the shape of the relationship between the factor scores and the anchor point estimates as the number of anchor point in the utilized Italian and western German data sets was small compared to the number of regions. The UK data set, where half of the regions are also anchor points, points at a linear relationship. The differences in target group, estimation technique, and referred time period of the UK anchor point estimates, however, are crucial limitations. The question of the appropriate function to link factor scores and anchor point estimates can be studied in more detail with an Austrian data set, where for each region a prevalence estimate with the target group of problematic opiate users is available (Uhl & Seidler, 2000).

The authors collected data on substitution therapies, police charges and drug-related deaths for all Austrian provinces for 1995 and calculated regional capture-recapture estimates. In the case of Vienna all three data sources were employed. Because of the low number of drug-related deaths in all the other regions, prevalence of problem opiate use was estimated by two-sample-capture-recapture with substitution therapies and police charges. All regional prevalence estimates sum up to 19,522 (see table 37). Furthermore, a three sample-capture-recapture estimate was calculated for all of Austria, resulting in 17,341 problem opiate users. A detailed description of model selection and its problems is given in Uhl & Seidler (2000).

In addition to the analysis of the relationship between factor scores and anchor point estimates, the data set suits for cross-validation of the capture-recapture method and the multivariate indicator method. The cross-validation was restricted to three anchor points, as on the one hand more than three anchor points are rarely available and on the other hand the application of only two anchor points yielded bad results in the sensitivity analysis of the UK data.

Furthermore, the variant “PCA with original data” was applied with the same sets of anchor points. The variants based on correlation of the indicator values or the indicator rates were not calculated as the correlations are extremely high: With regard to the original indicator values the correlations are 0.998 (substitution and police), 0.994 (substitution and deaths), and 0.991 (police and deaths). With regard to the indicator rates the corresponding correlations are 0.977 (substitution and police), 0.968 (substitution and deaths), and 0.982 (police and deaths).

Note, that similar results obtained by both methods does not mean that these estimates are similar to true prevalence. It is, however, reasonable, to treat the capture-recapture estimate as “golden standard” as much more information is used and the estimates are therefore expected to be superior to those obtained by the multivariate indicator method: In the case of nation-wide capture-recapture the overlap between the different indicators has to be known, which is not used in the case of the multivariate indicator method.

8.1 Austrian Data

The data set is given in table 37.

Table 37: Parameters and anchor points for the multivariate indicator method for Austria, 1995

Regions	Population 15-54 years	A	B	C	G anchor points
Burgenland	150,180	10	30	1	75
Carinthia	316,533	46	43	5	132
Lower Austria	841,527	184	234	8	1485
Salzburg	298,893	77	54	2	462
Styria	683,139	53	86	3	651
Tyrol	386,414	187	215	9	3093
Upper Austria	791,522	202	167	21	1874
Vienna	937,827	1653	1592	127	10953
Vorarlberg	202,260	210	277	22	797
Total	4569,295	2622	2698	198	19522

A Number of substitution therapies

B Police charges

C Drug-related deaths

G Estimated values of regional IDU population, independently obtained

8.2 PCA with indicator rates

8.2.1 Shape of the relationship between factor scores and anchor point estimates

Figure 23 shows the scatterplot of the factor scores obtained by PCA of the indicator rates and C-RC prevalence rates by regions. The regions Salzburg, Styria, Upper Austria, Lower Austria, and Vienna, i.e. 5 out of 9 regions, lie approximately on a straight line. Burgenland and Carinthia are close to this line. Tyrol and Vorarlberg are situated in approximately the same distance above respectively below this line. Altogether, the relationship between factor scores and C-RC prevalence rates seems to be linear.

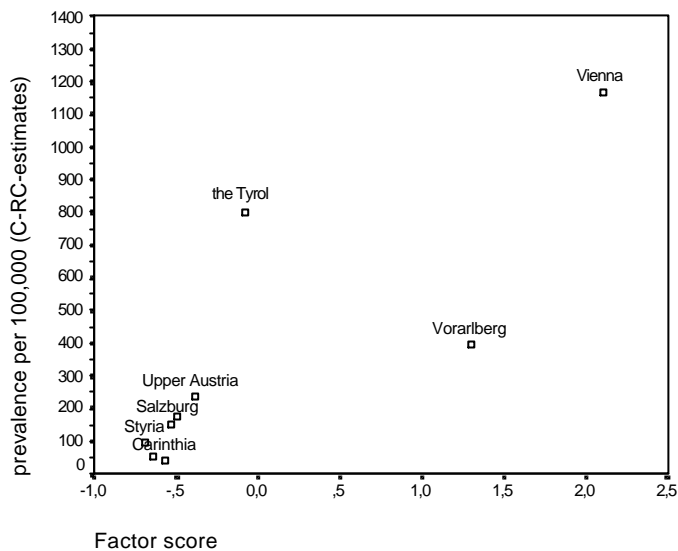


Figure 23: Scatterplot of factor scores and regional C-RC prevalence rates

8.2.2 Prevalence estimates

In table 38 prevalence estimates based on different sets of anchor points are compared to the unstratified C-RC estimate of all of Austria. The estimate employing Carinthia, Styria, and Tyrol deviates least from the reference value though the regression line through these anchor points is much steeper than the line connecting Salzburg, Styria, Upper Austria, Lower Austria, and Vienna. Employing three of these five regions (Lower Austria, Styria, Vienna) results in an estimate close to the sum of the regional C-RC-estimates. Using at least two of the regions with the highest prevalence rates – Tyrol, Vienna, Vorarlberg – as anchor points leads to large deviations from the reference value.

Table 38: Effects of different anchor points: changes of national prevalence of Austria

Anchor Points	Prevalence Austria	Difference
Capture-Recapture estimate of all of Austria	17,341	Reference
Tyrol, Vienna, Vorarlberg	30,946	13605
Carinthia, Styria, Tyrol	17,141	200
Lower Austria, Styria, Vienna	18,901	1560
Lower Austria, Vienna, Vorarlberg	13,472	-3869
Lower Austria, Tyrol, Vienna	27,379	10038

Table 39 shows the effects of the choice of different anchor points on the prevalence rates of the regions. Each column presents the results of the calculation of one model. Regions utilised as anchor points are marked with asterisk [*].

Table 39: Effects of different anchor points: Estimated prevalence rates of Austrian regions

Region	Prevalence rate					C-RC
	Multivariate indicator method					
Burgenland	873	94	174	49	533	75
Carinthia	1867	* 291	456	180	1195	132
Lower Austria	5044	1050	* 1475	* 713	* 3392	1485
Salzburg	1773	310	464	200	1156	462
Styria	3929	* 286	* 659	102	2315	651
Tyrol	* 2503	* 1123	1286	865	* 2052	3093
Upper Austria	4837	1307	1691	939	3437	1874
Vienna	* 8482	10968	* 10954	* 9025	* 11354	10953
Vorarlberg	* 1639	1713	1743	* 1398	1944	797

It can be seen that the prevalence rates in a row differ, i.e. they differ between the various models. Apparently, none of the analysed models yields regional estimates close to the C-RC estimates – even if the corresponding national estimate is close to the national C-RC-estimate. Thus, the multivariate indicator method seems to be inappropriate for regional prevalence estimation.

8.3 PCA with indicator values

8.3.1 Shape of the relationship between factor scores and anchor point estimates

Figure 24 depicts the scatterplot of the factor scores based on the original values and the regional C-RC-estimates. Except for Vienna, all regions are positioned close to a rather steep straight line. The highest prevalences are found in Vienna, followed by Tyrol and Upper Austria. Vorarlberg, which ranks third in the ordering of prevalence per 100,000 inhabitants ranks fifth in the ordering of prevalences. The least problem opiate users are found in Carinthia and Burgenland.

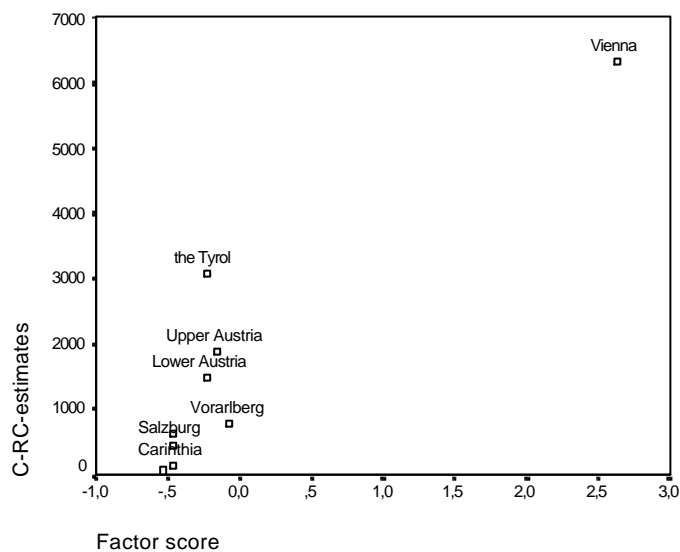


Figure 24: Scatterplot of factor scores and regional C-RC prevalence rates

8.3.2 Prevalence estimates

Table 40 shows the variations of the total prevalences dependent on the different models. The first column lists the utilised anchor points, the second presents the national prevalence and the third column shows the difference referring to the unstratified C-RC estimate of all of Austria.

Table 40: Effects of different anchor points: changes of national prevalence of Austria

Anchor Points	Prevalence Austria	Difference
Capture-Recapture estimate of all of Austria	17,341	Reference
Tyrol, Vienna, Vorarlberg	22,211	4,870
Carinthia, Styria, Tyrol	52,946	35,605
Lower Austria, Styria, Vienna	19,888	2,547
Lower Austria, Vienna, Vorarlberg	15,361	-1,980
Lower Austria, Tyrol, Vienna	26,918	9,577

As all regions with the exception of Vienna are positioned close to a steep line the multivariate indicator method without Vienna as anchor point leads to a high prevalence estimate – it is

actually nearly twice the reference value. The other estimates range from 15,361 to 26,918. The estimate 26.918 was calculated with the anchor points Lower Austria, Tyrol, and Vienna and is similar to the estimate (27,379) obtained by PCA per 100,000 inhabitants. Contrary to PCA per 100,000 inhabitants, PCA with the original values and the anchor points Tyrol, Vienna, and Vorarlberg yields an estimate that deviates only by 4,870 from the unstratified C-RC-estimate and by 2,689 from the stratified C-RC-estimate for all of Austria.

Table 41 shows the effects of the choice of different anchor points and the choice of a varying number of anchor points on the prevalence rates of the regions. Each column presents the results of the calculation of one model. Regions utilised as anchor points are marked with asterisk [*].

Table 41: Effects of different anchor points: prevalence estimates of Austrian regions

Region	Prevalence					C-RC
	Multivariate indicator method					
Burgenland	744	0	416	0	1358	75
Carinthia	953	* 286	633	19	1556	132
Lower Austria	1732	3135	* 1443	* 874	* 2293	1485
Salzburg	962	316	642	28	1563	462
Styria	1008	* 485	* 690	79	1607	651
Tyrol	* 1723	* 3105	1435	865	* 2285	3093
Upper Austria	1969	4005	1691	1135	2518	1874
Vienna	* 8482	10968	* 10954	* 9025	* 11354	10953
Vorarlberg	* 1639	1713	1743	* 1398	1944	797

As with PCA per 100,000 inhabitants, the regional estimates obtained by PCA with the original data vary to great extent between the different models. For example, in the case of Salzburg, the estimates calculated with the multivariate indicator method range from 28 to 1563 – whereas the C-RC-estimate was 462. Therefore, it has to be concluded that the multivariate indicator method is inappropriate for regional estimation.

9 Dissemination and/or exploitation of results

The project consolidated a multi-national, multi-disciplinary network of experts in drug use. The members of the network of the national prevalence estimation project were: Rita Augustin, Catherine Comiskey, Antonia Domingo, Martin Frischer, Petra Kümmler, Fabio Mariani, Carla Rossi, Alfred Uhl. The project was coordinated by Ludwig Kraus. Several meetings were held, with a varying group of participants. Jean-Michel Costes, as well as Francois Beck, Lucilla Rava, Katusca Berretta, and Lucas Wiessing were participating.

The output was disseminated in the form of meetings, publications, and the Internet (EMCDDA Internet site).

- EMCDDA (2000a). Final Report - Study to obtain Comparable National Estimates of Problem Drug Use Prevalence for all EU Member States. EMCDDA Project CT. 97.EP.04. Lisbon: EMCDDA.
- EMCDDA (2000b). Guidelines - Methods of national prevalence estimation. EMCDDA Project CT. 99.RTX.05. Lisbon: EMCDDA.
- Frischer, M., Hickman, M., Kraus, L., Mariani, F. & Wiessing, L. (2001): A comparison of different methods for estimating the prevalence of problematic drug misuse in Great Britain. *Addiction*, 96, 1465-1476.
- Kraus, L., Kümmler, P., Augustin, R., et al. "Estimating Prevalence of Problem Drug Use at National Level in Countries of the European Union and Norway". Publication. Completed
- Kümmler, P., Kraus, L.; Augustin, R. et al. „Estimating Prevalence of Problem Drug Use at National Level in Countries of the European Union“. Presentation; 11th Conference on the Reduction of Drug-related Harm; 10.04.2000; Jersey.

References

- Brugal, M. et al. (1999). A small area analysis estimating the prevalence of addiction to opioids in Barcelona, 1993. *Journal of Epidemiology & Community Health*, 53, 488-494.
- EMCDDA (1997a). *National prevalence estimates. Improvement of comparability of national estimates of addiction*. Final report. EMCDDA Project CT.96.EP.06.
- EMCDDA (1997b). *Methodological pilot study of local level prevalence estimates*. Final report. EMCDDA Project CT.96.EP.07.
- EMCDDA (1997c). *Annual report on the state of the drug problem in the European Union*. 1997. Lisbon: European Communities.
- EMCDDA (1998a). *Annual report on the state of the drug problem in the European Union*. 1998. Lisbon: European Communities.
- EMCDDA (1998b). *Methodological guidelines to estimate the prevalence of problem drug use on the local level*. EMCDDA Project CT.96.EP.06. Lisbon: EMCDDA.
- EMCDDA (2000a). Final Report - Study to obtain Comparable National Estimates of Problem Drug Use Prevalence for all EU Member States. EMCDDA Project CT. 97.EP.04. Lisbon: EMCDDA.
- Frischer, M., Hickman, M., Kraus, L., Mariani, F. & Wiessing, L. (2001). A comparison of different methods for estimating the prevalence of problematic drug misuse in Great Britain. *Addiction*, 96, 1465-1476.
- Mariani, F. (1999). Country report: Italy. In European Monitoring Centre for Drugs and Drug Addiction (EMCDDA). *Study to obtain comparable national estimates of problem drug use prevalence for all EU member States*. Lisbon: EMCDDA
- Person, P. H., Retka, R. L. & Woodward, J. A. (1977). *Toward a heroin problem index – an analytical model for drug use indicators*. Technical Paper. Rockville: National Institute on Drug Abuse.
- Person, P. H., Retka, R. L. & Woodward, J. A. (1978). *A method for estimating heroin use prevalence*. Technical Paper. Rockville: National Institute on Drug Abuse
- Sartor, F. & Walkiers, D. (2001). The prevalence of problematic drug use. Methodological aspects and feasibility in Belgium. *Arch Public Health*, 59, 77-100.
- Uhl, A. & Seidler, D. (2000). *Prevalence estimate of problematic opiate consumption in Austria*. Ludwig-Boltzmann Institute for Addiction Research: Vienna.
- Wickens, T. D. (1993). Quantitative methods for estimating the size of a drug-using population. *Journal of Drug Issues*, 23, 185-216.
- Woodward, J. R., Retka, R. L. & Ng, L. (1984). Construct validity of heroin abuse estimators. *International Journal of the Addictions*, 19, 93-117.

Appendix

In the following you find the SPSS syntax files for the six variants examined in the report. The data files contains of the variables REGION, A,B,C,D,E, F15_54, and G. A,...,E denote the regional indicator values, F15_54 the size of the regional 15-54-year-old population, and G the anchor point estimates. G is set to the system missing value if a region is no anchor point.

SPSS-Syntax of the variant "PCA with original data".

```
GET FILE "G:\PL-Kraus\50-454TS\Prog\Juni_01\UK.SAV".
```

```
*
*****
**
* Standardizing the indicator values.
* There is no SPSS-routine for standardizing variables. As calculating
mean and standard deviation for each variable and standardizing it after
having read its mean and standard deviation is very time-consuming we
applied a trick: We introduced a new variable named "one".The value of
this variable is 1 for each region. Taking the standardized residuals of
the linear regression of the
indicator values on "one" provides us with the standardized indicator
values.
* SPSS automatically names this standardized residuals zre_1, zre_2, etc.
```

```
COMPUTE one = 1 .
EXECUTE .
```

```
REGRESSION
/MISSING LISTWISE
/STATISTICS COEFF OUTS R ANOVA
/CRITERIA=PIN(.05) POUT(.10)
/ORIGIN
/DEPENDENT a
/METHOD=ENTER one
/SAVE ZRESID .
```

```
REGRESSION
/MISSING LISTWISE
/STATISTICS COEFF OUTS R ANOVA
/CRITERIA=PIN(.05) POUT(.10)
/ORIGIN
/DEPENDENT b
/METHOD=ENTER one
/SAVE ZRESID .
```

```
REGRESSION
/MISSING LISTWISE
/STATISTICS COEFF OUTS R ANOVA
/CRITERIA=PIN(.05) POUT(.10)
/ORIGIN
/DEPENDENT c
/METHOD=ENTER one
/SAVE ZRESID .
```

```
REGRESSION
/MISSING LISTWISE
/STATISTICS COEFF OUTS R ANOVA
/CRITERIA=PIN(.05) POUT(.10)
/ORIGIN
/DEPENDENT d
```

```

/METHOD=ENTER one
/SAVE ZRESID .

REGRESSION
/MISSING LISTWISE
/STATISTICS COEFF OUTS R ANOVA
/CRITERIA=PIN(.05) POUT(.10)
/ORIGIN
/DEPENDENT e
/METHOD=ENTER one
/SAVE ZRESID .

*
*****
**
* Principal components analysis.
* saving the estimated values of the first component (variable name
provided by SPSS: fac1_1).

FACTOR
/VARIABLES zre_1 to zre_5 /MISSING LISTWISE /ANALYSIS zre_1 to zre_5
/PRINT UNIVARIATE
/FORMAT SORT
/CRITERIA FACTORS(1) ITERATE(50)
/EXTRACTION PC
/ROTATION NOROTATE
/SAVE REG(ALL)
/METHOD=CORRELATION .
* *****
* Linear regression of g on fac1_1.
* The new variable pre_1 contains the regional prevalence population.

REGRESSION
/MISSING LISTWISE
/STATISTICS OUTS
/CRITERIA=PIN(.05) POUT(.10)
/NOORIGIN
/DEPENDENT g /METHOD=ENTER fac1_1
/SAVE PRED .

*****
* prevalence estimates for each region.

compute pre_rate=pre_1*100000/f15_54.
execute.
rename variables pre_1=preval.
Variable label preval "estimated prevalence, PCA with original data"
/pre_rate "prevalence rates per 100,000, PCA with original
data" .
LIST
VARIABLES=region g preval pre_rate.

* national prevalence estimate.
FREQUENCIES
VARIABLES=preval /FORMAT=NOTABLE
/STATISTICS=SUM

```

/ORDER= ANALYSIS .

SPSS-Syntax of the variant "PCA per 100,000".

```

GET FILE='C:\IFT\TSER\GERMANY\Germany.sav' .

* *****.
* calculation of the indicator rates per 100,000 inhabitants.

COMPUTE a_f = a*100000 / f15_54 .
COMPUTE b_f = b*100000 / f15_54 .
COMPUTE c_f = c*100000 / f15_54 .
COMPUTE d_f = d*100000 / f15_54 .
COMPUTE e_f = e*100000 / f15_54 .
COMPUTE g_f = g*100000 / f15_54 .

* *****.
* Standardizing the indicator rates.
* There is no SPSS-routine for standardizing variables. As calculating
mean and standard deviation for each variable and standardizing it after
having read its mean and standard deviation is very time-consuming we
applied a trick: We introduced a new variable named "one".The value of
this variable is 1 for each region. Taking the standardized residuals of
the linear regression of the indicator rates on "one" provides us with
the standardized indicator rates.
* SPSS automatically names this standardized residuals zre_1, zre_2, etc.

COMPUTE one = 1 .
EXECUTE .

REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS R ANOVA
  /CRITERIA=PIN(.05) POUT(.10)
  /ORIGIN
  /DEPENDENT a_f
  /METHOD=ENTER one
  /SAVE ZRESID .

REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS R ANOVA
  /CRITERIA=PIN(.05) POUT(.10)
  /ORIGIN
  /DEPENDENT b_f
  /METHOD=ENTER one
  /SAVE ZRESID .

REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS R ANOVA
  /CRITERIA=PIN(.05) POUT(.10)
  /ORIGIN
  /DEPENDENT c_f
  /METHOD=ENTER one

```

```

/SAVE ZRESID .

REGRESSION
/MISSING LISTWISE
/STATISTICS COEFF OUTS R ANOVA
/CRITERIA=PIN(.05) POUT(.10)
/ORIGIN
/DEPENDENT d_f
/METHOD=ENTER one
/SAVE ZRESID .

REGRESSION
/MISSING LISTWISE
/STATISTICS COEFF OUTS R ANOVA
/CRITERIA=PIN(.05) POUT(.10)
/ORIGIN
/DEPENDENT e_f
/METHOD=ENTER one
/SAVE ZRESID .

*
*****
**
* Principal components analysis.
* saving the estimated values of the first component (variable name
provided by SPSS: fac1_1).

FACTOR
/VARIABLES zre_1 zre_2 zre_3 zre_4 zre_5 /MISSING LISTWISE /ANALYSIS
zre_1
zre_2 zre_3 zre_4 zre_5
/PRINT UNIVARIATE INITIAL EXTRACTION
/FORMAT SORT
/CRITERIA FACTORS(1) ITERATE(50)
/EXTRACTION PC
/ROTATION NOROTATE
/SAVE REG(ALL)
/METHOD=CORRELATION .

*****
*****
* Linear regression of g_f on fac1_1.
* The new variable pre_1 contains the population estimates per 100,000
inhabitants.

REGRESSION
/MISSING LISTWISE
/STATISTICS COEFF OUTS R ANOVA
/CRITERIA=PIN(.05) POUT(.10)
/NOORIGIN
/DEPENDENT g_f
/METHOD=ENTER fac1_1
/SAVE PRED .

*****
* prevalence estimates for each region.

```

PART 2 – NATIONAL PREVALENCE - 83

```
compute preval=pre_1*f15_54/100000.
execute.
rename variables pre_1=pre_rate.
Variable label preval "estimated prevalence, PCA per 100,000"
                /pre_rate "prevalence rates per 100,000, PCA per 100,000" .
LIST
  VARIABLES=region g preval pre_rate.

* national prevalence estimate.
FREQUENCIES
  VARIABLES=preval /FORMAT=NOTABLE
  /STATISTICS=SUM
  /ORDER= ANALYSIS .
```

SPSS-Syntax of the variant "PCA with ranks of the original data".

```
GET FILE "G:\PL-Kraus\50-454TS\Prog\Juni_01\UK.SAV".
```

```
* *****.
* Assigning the ranks to the variables a,...,e.
* variable names: ra,...,re.
```

```
RANK
```

```
VARIABLES=a b c d e (A) /RANK /PRINT=NO /TIES=MEAN .
*****.
*****.
* Standardizing the indicator values.
* There is no SPSS-routine for standardizing variables. As calculating
mean and standard deviation for each variable and standardizing it after
having read its mean and standard deviation is very time-consuming we
applied a trick: We introduced a new variable named "one".The value of
this variable is 1 for each region. Taking the standardized residuals of
the linear regression of the indicator values on "one" provides us with
the standardized indicator values.
* SPSS automatically names this standardized residuals zre_1, zre_2, etc.
```

```
COMPUTE one = 1 .
EXECUTE .
```

```
REGRESSION
```

```
/MISSING LISTWISE
/STATISTICS COEFF OUTS R ANOVA
/CRITERIA=PIN(.05) POUT(.10)
/ORIGIN
/DEPENDENT ra
/METHOD=ENTER one
/SAVE ZRESID .
```

```
REGRESSION
```

```
/MISSING LISTWISE
/STATISTICS COEFF OUTS R ANOVA
/CRITERIA=PIN(.05) POUT(.10)
/ORIGIN
/DEPENDENT rb
/METHOD=ENTER one
/SAVE ZRESID .
```

```
REGRESSION
```

```
/MISSING LISTWISE
/STATISTICS COEFF OUTS R ANOVA
/CRITERIA=PIN(.05) POUT(.10)
/ORIGIN
/DEPENDENT rc
/METHOD=ENTER one
/SAVE ZRESID .
```

```
REGRESSION
```

```
/MISSING LISTWISE
/STATISTICS COEFF OUTS R ANOVA
```

```

/CRITERIA=PIN(.05) POUT(.10)
/ORIGIN
/DEPENDENT rd
/METHOD=ENTER one
/SAVE ZRESID .

REGRESSION
/MISSING LISTWISE
/STATISTICS COEFF OUTS R ANOVA
/CRITERIA=PIN(.05) POUT(.10)
/ORIGIN
/DEPENDENT re
/METHOD=ENTER one
/SAVE ZRESID .

*
*****
**
* Principal components analysis.
* saving the estimated values of the first component (variable name
provided by SPSS: fac1_1).

FACTOR
/VARIABLES zre_1 to zre_5 /MISSING LISTWISE /ANALYSIS zre_1 to zre_5
/PRINT UNIVARIATE
/FORMAT SORT
/CRITERIA FACTORS(1) ITERATE(50)
/EXTRACTION PC
/ROTATION NOROTATE
/SAVE REG(ALL)
/METHOD=CORRELATION .
* *****
* Linear regression of g on fac1_1.
* The new variable pre_1 contains the regional prevalence population.

REGRESSION
/MISSING LISTWISE
/STATISTICS OUTS
/CRITERIA=PIN(.05) POUT(.10)
/NOORIGIN
/DEPENDENT g /METHOD=ENTER fac1_1
/SAVE PRED .

*****
* prevalence estimates for each region.

compute pre_rate=pre_1*100000/f15_54.
execute.
rename variables pre_1=preval.
Variable label preval "estimated prevalence, PCA with ranks of original
data"
          /pre_rate "prevalence rates per 100,000, PCA with ranks of
original data" .
LIST
  VARIABLES=region g preval pre_rate.

```

```
* national prevalence estimate.  
FREQUENCIES  
  VARIABLES=preval  /FORMAT=NOTABLE  
  /STATISTICS=SUM  
  /ORDER= ANALYSIS .
```

SPSS-Syntax of the variant "PCA with ranks per 100,000".

```

GET
  FILE='C:\IFT\TSER\GERMANY\Germany.sav' .

*
*****
**
* calculation of the indicator rates per 100,000 inhabitants.

COMPUTE a_f = a*100000 / f .
COMPUTE b_f = b*100000 / f .
COMPUTE c_f = c*100000 / f .
COMPUTE d_f = d*100000 / f .
COMPUTE e_f = e*100000 / f .
COMPUTE g_f = g*100000 / f .
execute.

* *****
* Assigning the ranks to the variables a_f,...,e_f.
* variable names: ra_f,...,re_f.

RANK
  VARIABLES=a_f b_f c_f d_f e_f (A) /RANK /PRINT=NO /TIES=MEAN .

*****
*****
* Standardizing the ranks of the indicator rates.
* There is no SPSS-routine for standardizing variables. As calculating
mean and standard deviation for each variable and standardizing it after
having read its mean and standard deviation is very time-consuming we
applied a trick: We intro-duced a new variable named "one".The value of
this variable is 1 for each region. Taking the standardized residuals of
the linear regression of the ranks of the indicator rates on "one"
provides us with the needed standardized values.
* SPSS automatically names this standardized residuals zre_1, zre_2, etc.

REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS R ANOVA
  /CRITERIA=PIN(.05) POUT(.10)
  /ORIGIN
  /DEPENDENT ra_f
  /METHOD=ENTER one
  /SAVE ZRESID .

REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS R ANOVA
  /CRITERIA=PIN(.05) POUT(.10)
  /ORIGIN
  /DEPENDENT rb_f
  /METHOD=ENTER one
  /SAVE ZRESID .

```

REGRESSION

```

/MISSING LISTWISE
/STATISTICS COEFF OUTS R ANOVA
/CRITERIA=PIN(.05) POUT(.10)
/ORIGIN
/DEPENDENT rc_f
/METHOD=ENTER one
/SAVE ZRESID .

```

REGRESSION

```

/MISSING LISTWISE
/STATISTICS COEFF OUTS R ANOVA
/CRITERIA=PIN(.05) POUT(.10)
/ORIGIN
/DEPENDENT rd_f
/METHOD=ENTER one
/SAVE ZRESID .

```

REGRESSION

```

/MISSING LISTWISE
/STATISTICS COEFF OUTS R ANOVA
/CRITERIA=PIN(.05) POUT(.10)
/ORIGIN
/DEPENDENT re_f
/METHOD=ENTER one
/SAVE ZRESID .

```

*

**.

* Principal components analysis.

* saving the estimated values of the first component (variable name provided by SPSS: fac1_1).

FACTOR

```

/VARIABLES zre_1 to zre_5 /MISSING LISTWISE /ANALYSIS zre_1 to zre_5
/PRINT UNIVARIATE
/FORMAT SORT
/CRITERIA FACTORS(1) ITERATE(50)
/EXTRACTION PC
/ROTATION NOROTATE
/SAVE REG(ALL)
/METHOD=CORRELATION .

```

*****.

* Linear regression of g_f on fac1_1.

* The new variable pre_1 contains the population estimates per 100,000 inhabitants.

REGRESSION

```

/MISSING LISTWISE
/STATISTICS COEFF OUTS R ANOVA
/CRITERIA=PIN(.05) POUT(.10)
/NOORIGIN

```

PART 2 – NATIONAL PREVALENCE - 89

```
/DEPENDENT g_f
/METHOD=ENTER fac1_1
/SAVE PRED .

*****.
* prevalence estimates for each region.

compute preval=pre_1*f15_54/100000.
execute.
rename variables pre_1=pre_rate.
Variable label preval "estimated prevalence, PCA with ranks per 100,000"
      /pre_rate "prevalence rates per 100,000, PCA with ranks per
100,000" .
LIST
  VARIABLES=region g preval pre_rate.

* national prevalence estimate.
FREQUENCIES
  VARIABLES=preval /FORMAT=NOTABLE
  /STATISTICS=SUM
  /ORDER= ANALYSIS .
```



```
REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS R ANOVA
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT g
  /METHOD=ENTER a e
  /SAVE PRED .
```

```
REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS R ANOVA
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT g
  /METHOD=ENTER b c
  /SAVE PRED .
```

```
REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS R ANOVA
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT g
  /METHOD=ENTER b d
  /SAVE PRED .
```

```
REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS R ANOVA
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT g
  /METHOD=ENTER b e
  /SAVE PRED .
```

```
REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS R ANOVA
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT g
  /METHOD=ENTER c d
  /SAVE PRED .
```

```
REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS R ANOVA
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT g
  /METHOD=ENTER c e
  /SAVE PRED .
```

```
REGRESSION
```

```

/MISSING LISTWISE
/STATISTICS COEFF OUTS R ANOVA
/CRITERIA=PIN(.05) POUT(.10)
/NOORIGIN
/DEPENDENT g
/METHOD=ENTER d e
/SAVE PRED .

* *****.
* Selection of the appropriate regression by comparison with the variable
coll (i.e. the former variable aux1).
do if (coll=1).
compute preval=pre_1.
else if (coll=2).
compute preval=pre_2.
else if (coll=3).
compute preval=pre_3.
else if (coll=4).
compute preval=pre_4.
else if (coll=5).
compute preval=pre_5.
else if (coll=6).
compute preval=pre_6.
else if (coll=7).
compute preval=pre_7.
else if (coll=8).
compute preval=pre_8.
else if (coll=9).
compute preval=pre_9.
else.
compute preval=pre_10.
end if.
execute.

*
*****
.
* prevalence estimates for each region.

compute pre_rate=preval*100000/f15_54.
execute.
Variable label preval "estimated prevalence, Correlation with original
data"
      /pre_rate "prevalence rates per 100,000, Correlation with original
data" .
LIST
  VARIABLES=region g preval pre_rate.

* national prevalence estimate.
FREQUENCIES
  VARIABLES=preval /FORMAT=NOTABLE
  /STATISTICS=SUM
  /ORDER= ANALYSIS .

```

SPSS-Syntax of the variant "Correlation per 100,000".

```

GET
  FILE='G:\PL-Kraus\50-454TS\Prog\nov_01\Italy\Italy_95.sav'.

*
*****
**
* calculation of the indicator rates per 100,000 inhabitants.

COMPUTE a_f = a*100000 / f15_54.
COMPUTE b_f = b*100000 / f15_54.
COMPUTE c_f = c*100000 / f15_54.
COMPUTE d_f = d*100000 / f15_54.
COMPUTE e_f = e*100000 / f15_54.
COMPUTE g_f = g*100000 / f15_54.
execute.

* *****
* Here: selection of the two indicator rates with the smallest absolute
value of the correlation coefficient.
* The result is saved in the file matrix.sav. This file contains 12 rows
and 7 columns in the case of 5 variables a_f,...,e_f.* 1st row: Means,
2nd row: Standard deviations, 3rd-7th row: Sample sizes, 8th-12th row:
Correlation coefficients.
* 1st column: Row type, 2nd column: Variable names, 3rd-7th column:
Variables a_f,...,e_f.

CORRELATIONS
/VARIABLES=a_f b_f c_f d_f e_f
/PRINT=TWOTAIL NOSIG
/MISSING=PAIRWISE
/Matrix=out("C:\IFT\TSER\allgemein\matrix.sav").

* Reading the file matrix.sav.
* Saving the correlation coefficients to the matrix MIM1.
* Saving the smallest absolute value of the correlation coefficients to
the auxiliary variable aux.

matrix.
get MIM
  /file="C:\IFT\TSER\allgemein\matrix.sav".
compute MIM1=MIM(8:12,3:7).
compute aux=MMin(abs(MIM1)).

* Comparing all correlation coefficients with aux.
* Building a new variable aux1 that contains the position of the smallest
absolute value of the correlations coefficients in the correlation
matrix.
* aux1 is a vector, the length of the vector is the number of regions.
* If e.g. the smallest absolute value of the correlations coefficients is
the correlation coefficient of a_f and b_f the vector {1;...;1} will be
assigned to aux1.

```



```
/NOORIGIN  
/DEPENDENT g_f  
/METHOD=ENTER a_f c_f  
/SAVE PRED .
```

```
REGRESSION  
/MISSING LISTWISE  
/STATISTICS COEFF OUTS R ANOVA  
/CRITERIA=PIN(.05) POUT(.10)  
/NOORIGIN  
/DEPENDENT g_f  
/METHOD=ENTER a_f d_f  
/SAVE PRED .
```

```
REGRESSION  
/MISSING LISTWISE  
/STATISTICS COEFF OUTS R ANOVA  
/CRITERIA=PIN(.05) POUT(.10)  
/NOORIGIN  
/DEPENDENT g_f  
/METHOD=ENTER a_f e_f  
/SAVE PRED .
```

```
REGRESSION  
/MISSING LISTWISE  
/STATISTICS COEFF OUTS R ANOVA  
/CRITERIA=PIN(.05) POUT(.10)  
/NOORIGIN  
/DEPENDENT g_f  
/METHOD=ENTER b_f c_f  
/SAVE PRED .
```

```
REGRESSION  
/MISSING LISTWISE  
/STATISTICS COEFF OUTS R ANOVA  
/CRITERIA=PIN(.05) POUT(.10)  
/NOORIGIN  
/DEPENDENT g_f  
/METHOD=ENTER b_f d_f  
/SAVE PRED .
```

```
REGRESSION  
/MISSING LISTWISE  
/STATISTICS COEFF OUTS R ANOVA  
/CRITERIA=PIN(.05) POUT(.10)  
/NOORIGIN  
/DEPENDENT g_f  
/METHOD=ENTER b_f e_f  
/SAVE PRED .
```

```
REGRESSION  
/MISSING LISTWISE  
/STATISTICS COEFF OUTS R ANOVA  
/CRITERIA=PIN(.05) POUT(.10)  
/NOORIGIN  
/DEPENDENT g_f
```

```

/METHOD=ENTER c_f d_f
/SAVE PRED .

REGRESSION
/MISSING LISTWISE
/STATISTICS COEFF OUTS R ANOVA
/CRITERIA=PIN(.05) POUT(.10)
/NOORIGIN
/DEPENDENT g_f
/METHOD=ENTER c_f e_f
/SAVE PRED .

REGRESSION

/MISSING LISTWISE
/STATISTICS COEFF OUTS R ANOVA
/CRITERIA=PIN(.05) POUT(.10)
/NOORIGIN
/DEPENDENT g_f
/METHOD=ENTER d_f e_f
/SAVE PRED .

* *****.
* Selection of the appropriate regression by comparison with the variable
coll (i.e. the former variable aux1).
do if (coll=1).
compute pre_rate=pre_1.
else if (coll=2).
compute pre_rate=pre_2.
else if (coll=3).
compute pre_rate=pre_3.
else if (coll=4).
compute pre_rate=pre_4.
else if (coll=5).
compute pre_rate=pre_5.
else if (coll=6).
compute pre_rate=pre_6.
else if (coll=7).
compute pre_rate=pre_7.
else if (coll=8).
compute pre_rate=pre_8.
else if (coll=9).
compute pre_rate=pre_9.
else.
compute pre_rate=pre_10.
end if.
execute.

*
*****
.
* prevalence estimates for each region.

compute preval=pre_rate*f15_54/100000.
execute.
Variable label preval "estimated prevalence, Correlation per 100,000"

```

PART 2 – NATIONAL PREVALENCE - 98

```

/pre_rate "prevalence rates per 100,000, Correlation per
100,000" .
LIST
  VARIABLES=region g preval pre_rate.

* national prevalence estimate.
FREQUENCIES
  VARIABLES=preval /FORMAT=NOTABLE
  /STATISTICS=SUM
  /ORDER= ANALYSIS .

```

- ¹ Person, P. H., Retka, R. L. & Woodward, J. A. (1976). *Toward a heroin problem index - an analytical model for drug use indicators*. Technical Paper. Rockville: National Institute on Drug Abuse.
- ² Person, P. H., Retka, R. L. & Woodward, J. A. (1977). *A method for estimating heroin use prevalence*. Technical Paper. Rockville: National Institute on Drug Abuse.

Table 1: Data and sources for the multiplier method

Parameter	Data	Source of information
n	Number of problem drug users who underwent treatment in a given year	Treatment centres
f	Probability for a problem drug user to be treated (in-treatment rate)	
n1	Number of registered problem drug users in a given year	Police
f1	Probability for a problem drug user to be registered by the police in that year	
n2	Number of problem drug users registered by the police for the first time (over a period reflecting mean duration of addiction)	
f2	Proportion of drug-related deaths that have previously been registered by the police as problem drug users (also over the same period)	
n	Number of drug-related deaths in a given year	Mortality
f	Probability of death among problem drug users in the same year	
n	Number of drug injectors being HIV positive in a given year	HIV
f	Probability of HIV positive drug injectors in the same year	