Estimating the prevalence of people who inject drugs in France, 2014

Eric Janssen - General Population Surveys Unit
French Monitoring Centre on Drugs and Drug Addictions - OFDT

PDU expert meeting

June 2017

Background

- Latest available estimates on IDU in France: 2006.
- Last month use only: n*=81000, e.g. 2.2‰ (15-64 year-olds).
- Overall number of users
- Need for updated, more detailed information to model the spreading of virus-borne diseases

Challenges (common)

- IDUs constitute an hard-to-reach population
- Escape traditional samplings. GPS provide poor results. Indirect methods favoured
- Capture-recapture using several data sources has gained considerable interest in the past decades: 2, 3, up to 4 sources (King et al)

Challenges (specific)

Several data sources or surveys in France, but:

- **Quality issues - Several data sources is no panacea: quality issues on data sources: geographical coverage, time span, definitions, etc.**

Data linkage is unlikely:

- **Counter effect of the generalization of databases: legal restrictions to protecting anonymity of illicit substance users in France**

- **Medical and field workers community in France reluctant to share and cross-check: anonymity as a essential pre-condition in reaching as many users as possible**

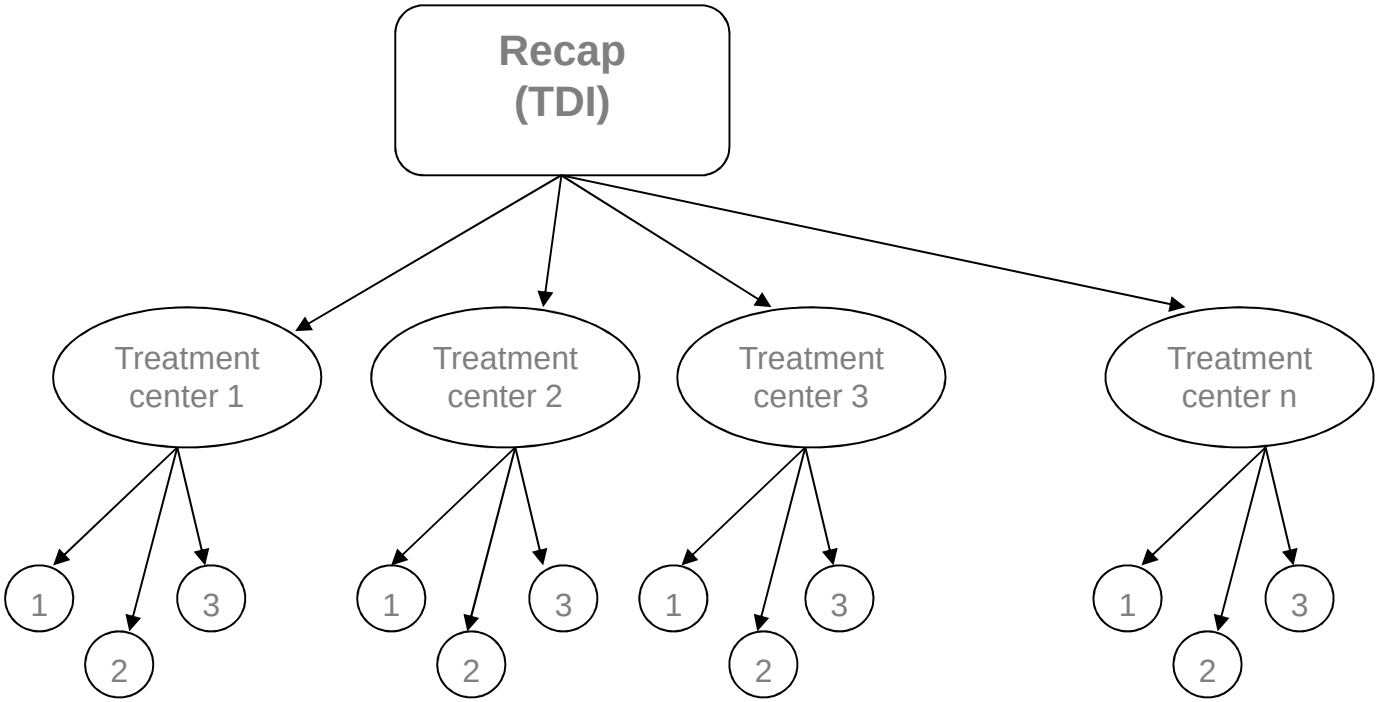
Proposal

- Capture-recapture – best method for detailed information
- Single source – more reliable

Data

- Extracted from RECAP, domestic application of TDI
- Compendium updated yearly since 2006
- Follow-up of patients full civil year, N>169000 in 2014
- Standardized questionnaire, relying on a conservative use (i.e. has remained stable since the study was launched)
- Information includes substance uses, route of administration, sociodemographics (gender, age, housing, households, incomes)
- Personal unique identifier provided by each TC (n=257 in metropolitan France)

Data (cont'd)


Recap
(TDI)
Treatment
center 1
Treatment
center 2
Treatment
center 3
Treatment
center n
1
2
3
1
2
3
1
2
3
1
2
3

Method

- Single-source CR gained popularity in the 1980's. Seminal work for Chao and Zelterman focusing on individuals observed/recorded once and twice

- More recently, extension of the Zelterman indicator to a Zelterman regression (Böhnind & van der heijden 2009) in order to capture heterogeneity at individual level (covariates) => uncontrolled heterogeneity yields biased (underestimated) figures

- Successfully applied to estimate PDU in Amsterdam (van der Heijden et al, 2013)

- Method applicable to restricted area. Way to expand it at national level?

Statistical framework

- Single source
- Zelterman indicator (1988) based on individuals recorded once or twice
- Zelterman regression (Böhning & van der Heijden 2009)
- Hierarchical structure: patients are nested within treatment centres = > Multilevel Zelterman regression

Results

- Last month estimated number of IDU: 86000 (69000; 110000) in 2014 => stable compared to 2006 (n=81000)
- Last year estimated number of IDU: 104000 (85000; 130000)
- Last year male IDUs: 80000 (61000; 106000)
- Last year female IDUs: 24000 (16000; 39000)
- 3 to 1 ratio of male to female observed in many other European countries

Results (cont'd)

	Last month IDU		
	Overall	Males	Females
Prevalence	2.2	3.2	1.0
95% CI	1.7; 2.7	2.5; 4.5	0.6; 1.6
	Last year IDU		
	Overall	Males	Females
Prevalence	2.6	4.0	1.2
95% CI	2.1; 3.2	3.1; 5.3	0.8; 1.9

- **Prevalence: ‰ of 15-64 year-olds. Source: RECAP 2014**

Results (cont'd)

	Last month IDU			
	Overall	**15-29 y-o**	**30-44 y-o**	**45-64 y-o**
Prevalence	2.2	2.7	2.9	1.2
95% CI	1.7; 2.7	2.4; 3.2	2.5; 3.3	0.9; 1.5
	Last year IDU			
	Overall	**15-29 y-o**	**30-44 y-o**	**45-64 y-o**
Prevalence	2.2	3.4	3.2	1.5
95% CI	2.1; 3.2	2.7; 4.4	2.7; 4.0	1.2; 1.8

- **Prevalence: 15-64 year-olds. Source: RECAP 2014**

Conclusion

- Reliable alternative method to provide estimates, controlling for heterogeneity at several level
- Here, two levels considered: treatment centres and individual. A third level (regions), whose gain proved marginal => restricted to simpler model
- Method applied to estimate the number of heroin users

Conclusion (cnt'd)

- Need for external assessment to validate estimates
- Method relies on a fair number of TC
- Further research to provide estimates at regional level
- Multisource approach needed – qualitative, ethnographic studies at local level. Injection not restricted to heroin users (smoke and snort): buprenorphin (old), most deprived stimulants users (new)
- TC: partial coverage stimulants uses => either alternative data base or another method

References

- Böhning & van der Heijden (2009), A covariate adjustment for zero-truncated approaches to estimating the size of hidden and elusive populations, Annals of Applied Statistics, 3: 595-610

- Janssen (2017), Estimating the number of people who inject drugs: A proposal to provide figures nationwide and its application to France, Journal of Public Health, doi: 10.1093/pubmed/fdx059

- King et al (2014), Estimating prevalence of injecting drug users and associated heroin-related death rates in England by using regional data and incorporating prior information, Journal of the Royal Statistical Society (A), 177: 209-236

- van der Heijden et al (2013), Methods for population size estimation of problem drug users using a single registration, International Journal of Drug Policy, 24: 614-618

- Zelterman D (1988), Robust estimation in truncated discrete distribution with application to capture-recapture experiments, Journal of Statistical Planning and Inference, 18: 225-237

Thank you for your attention

Zelterman indicator

$N^{estimated}=N^{observed}/(1-\exp(-2f_2/f_1))$

Where

$N^{observed}$=the total number of observed or recorded individuals,

f_1=individuals appearing once

f_2=individuals appearing twice

Zelterman regression

$$N^{estimated} = \sum_{i=1}^{N} \frac{I_i}{1 - \exp(-2\exp(\lambda_i))}$$

Where Ii=1 if individual i is recorded, 0 otherwise and λ is estimated by means of a logistic regression:

$$\lambda_i = Logit[Pr(y_i=1|X_i)] = \beta_0 + \beta_i X'_i + \varepsilon_i$$

Here, y is the dependent variable (coded 0 if a client is recorded once, coded 1 if recorded twice), X the matrix of independent variables, εi is an individual random error term, $\varepsilon \sim N(0;\sigma^2\varepsilon)$

Multilevel Zelterman regression

$\lambda_{ij}=\text{Logit}[Pr(y_{ij}=1|X_{ij}, u_{.j})]=\beta_0+\beta_{ij}X'_{ij}+u_{.j}+\varepsilon_{ij}$

where

$u_{.j}$ is the error associated with centre j, $u_{.j} \sim N(0; \sigma^2_{u.j})$

ε_{ij} the error associated with individual i counted in centre j, of null mean and a variance of σ^2