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
Background

- 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 precondition 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
1 2 3

Treatment center 2
1 2 3

Treatment center 3
1 2 3

Treatment center n
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

<table>
<thead>
<tr>
<th></th>
<th>Overall</th>
<th>Males</th>
<th>Females</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Prevalence</strong></td>
<td>2.2</td>
<td>3.2</td>
<td>1.0</td>
</tr>
<tr>
<td><strong>95% CI</strong></td>
<td>1.7; 2.7</td>
<td>2.5; 4.5</td>
<td>0.6; 1.6</td>
</tr>
</tbody>
</table>

### Last year IDU

<table>
<thead>
<tr>
<th></th>
<th>Overall</th>
<th>Males</th>
<th>Females</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Prevalence</strong></td>
<td>2.6</td>
<td>4.0</td>
<td>1.2</td>
</tr>
<tr>
<td><strong>95% CI</strong></td>
<td>2.1; 3.2</td>
<td>3.1; 5.3</td>
<td>0.8; 1.9</td>
</tr>
</tbody>
</table>

- Prevalence: % of 15-64 year-olds. Source: RECAP 2014
## Results (cont’d)

### Last month IDU

<table>
<thead>
<tr>
<th></th>
<th>Overall</th>
<th>15-29 y-o</th>
<th>30-44 y-o</th>
<th>45-64 y-o</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prevalence</td>
<td>2.2</td>
<td>2.7</td>
<td>2.9</td>
<td>1.2</td>
</tr>
<tr>
<td>95% CI</td>
<td>1.7; 2.7</td>
<td>2.4; 3.2</td>
<td>2.5; 3.3</td>
<td>0.9; 1.5</td>
</tr>
</tbody>
</table>

### Last year IDU

<table>
<thead>
<tr>
<th></th>
<th>Overall</th>
<th>15-29 y-o</th>
<th>30-44 y-o</th>
<th>45-64 y-o</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prevalence</td>
<td>2.2</td>
<td>3.4</td>
<td>3.2</td>
<td>1.5</td>
</tr>
<tr>
<td>95% CI</td>
<td>2.1; 3.2</td>
<td>2.7; 4.4</td>
<td>2.7; 4.0</td>
<td>1.2; 1.8</td>
</tr>
</tbody>
</table>

- 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


Thank you for your attention
Zelterman indicator

\[ N_{\text{estimated}} = \frac{N_{\text{observed}}}{1 - \exp(-2 \frac{f_2}{f_1})} \]

Where

- \( N_{\text{observed}} \) = the total number of observed or recorded individuals,
- \( f_1 \) = individuals appearing once
- \( f_2 \) = individuals appearing twice
Zelterman regression

\[ N_{\text{estimated}} = \sum_{i=1}^{N} \frac{1}{1 - \exp(-2\exp(\lambda_i))} \]

Where \( i_i = 1 \) if individual \( i \) is recorded, 0 otherwise and \( \lambda \) is estimated by means of a logistic regression:

\[ \lambda_i = \text{Logit}[\Pr(y_i = 1 | X_i)] = \beta_0 + \beta_i X_i' + \epsilon_i \]

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, \( \epsilon_i \) is an individual random error term, \( \epsilon \sim N(0; \sigma^2\epsilon) \)
Multilevel Zelterman regression

\[ \lambda_{ij} = \text{Logit}[\Pr(y_{ij} = 1 | X_{ij}, u_j)] = \beta_0 + \beta_{ij} X'_{ij} + u_j + \epsilon_{ij} \]

where

\( u_j \) is the error associated with centre \( j \), \( u_j \sim N(0; \sigma_{u.j}^2) \)

\( \epsilon_{ij} \) the error associated with individual \( i \) counted in centre \( j \), of null mean and a variance of \( \sigma^2 \)