Collecting DRID data in bio-behavioural surveys – EMCDDA toolkit methodology

MAGDALENA ROSINSKA, NATIONAL INSTITUTE OF PUBLIC HEALTH – NATIONAL INSTITUTE OF HYGIENE, WARSAW, POLAND

Outline

Bio – behavioural surveys as surveillance tool

Sampling and representativeness

Sampling methods in populations of PWD/PWID

Surveillance - concept

‘the on-going, systematic

collection, analysis, interpretation, and dissemination of data regarding a health-related event

for use in public health action to reduce morbidity and mortality and to improve health.

Data disseminated by a public health surveillance system can be used for immediate public health action, programme planning and evaluation, and formulating research hypotheses’

Drug related infectious diseases surveillance: blood-borne infections

Numerous data sources may be of use.

Case-based surveillance and secondary data usually not sufficient to fully understand the burden of illness (prevalence) and extisting behaviours.

Need to actively collect additional information.

Need to interpret available information together.

Proposed simplified criteria for surveillance studies

1. Operational simplicity and reasonable cost
   - the system should be sustainable within public health structure
2. Picking up new trends
   - reproducibility over time implies that detected changes reflect trends in population
3. Validity of information
   - representativeness of the sample and valid measurement

Studies for DRID surveillance

Repeated cross – sectional bio-behavioural surveys

Local context

- What are the data gaps
- What are the key indicators for designing and evaluating interventions
- What are the exising resources – infrastructure helpful in implementation, existing data systems
- What is the epidemiological situation

Decision on target population, methodology, frequency of surveys and geographical coverage.

Cross-sectional study

Describe target population at a single point in time.

Usually qunatitative information

Can measure both exposure and outcome (e.g. risky behaviour and infection) at the same time

- We do not know, which was first – the exposure or the outcome

Can measure biological indicators (i.e. laboratory results from biological samples), behavioural indicators (i.e. questionnaire) or both.

We cannot measure the whole population:

- Rely on taking a representative sample of the population

Sampling and representativeness

Representative sample

Warrants unbiased estimates of our indicators for the target population

- i.e. error is random

Results can be generalized to the target population

Requires **probablity sampling scheme**

- each individual of the target population can be sampled
- Probability of sampling is known for each individual

Non-probability sampling

Sampling methods that do not let us know in advance the likelihood of selecting for the sample each element or case from a population

Limited generalizability—one cannot judge representativeness.

Why use nonprobability samples?

- No list of target population members to sample from, target population difficult to reach
- Quick, convenient
- Well-suited for exploratory research

Probability vs non-probability sampling

Issue	Random sample	Non-random sample
Prone to selection bias	No	yes
Can generalize results to survey population	Yes	No
Can estimate precision of survey estimates	Yes	No
Results considered credible	Yes	No
Requires sampling frame	Yes	No
Requiring following fixed procedures	Yes	No
Method replicable	Yes	No

Designing a sampling scheme

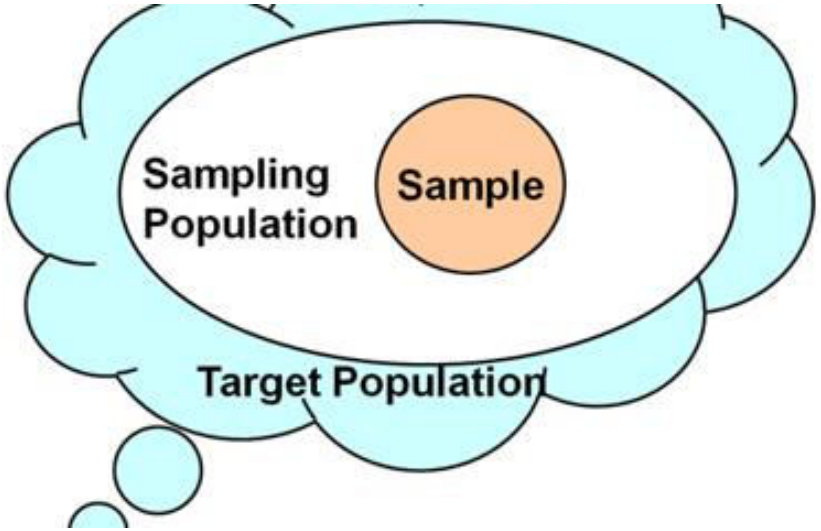
Define the target population

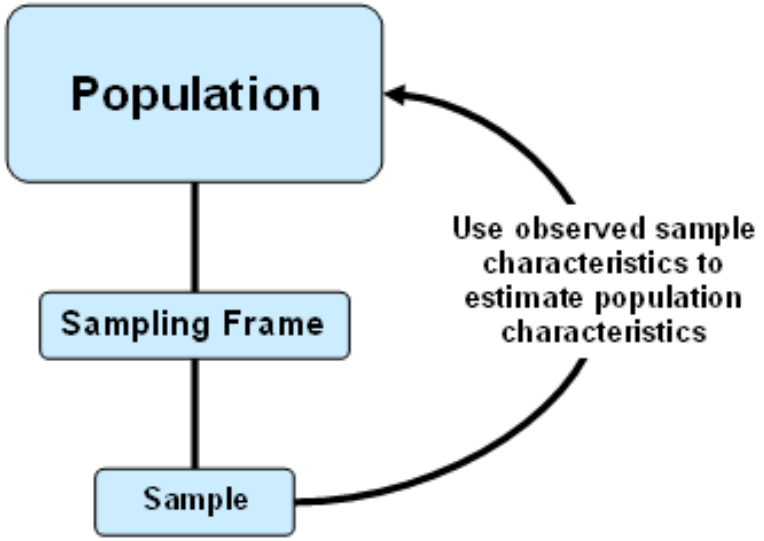
Establish the sampling frame

Take a sample from the sampling frame

Target population and sampling frame(s)

Sampling frame is the actual population from which the sample is drawn. Ideally the sampling frame is the target population or representative for the target population





Sampling methods for PWID

Selecting the target population

Injecting posed the highest risk for blood-borne infections -> monitor injecting population

A. Ever-IDUs who are also recent (last 12 months) problem drug users (PDUs).
B. Recent/current/active (last 4 weeks) IDUs.
C. Ever-IDUs in the general population (includes ever-IDUs who are not recent PDUs).
D. Recent (last 12 months) PDUs — always distinguishing ever- and never-IDUs.

Sampling frame(s)

Drug users remain a hard to reach, sometimes hidden population

In general no list to sample from, except for some subpopulations: PWID in drug treatment, PWID in prisons etc.

➔random sampling impossible

➔generalization of results of one survey among PWID is always problematic.

A possible solution is therefore to use different sampling frames in one study (e.g. recruitment from services and on the streets) in order to reduce overall bias when generalizing to the full PWID population or to use a method that approaches random sampling.

Sampling frame(s)

A. Community: PWID who are network with other PWID
B. Clients of out-patient services: PWID in contact with high or low threshold services for drug users, testing sites, social services
C. Closed setting patients: PWID in detox/rehab centers
D. Prisoners

Community is likely to be the closest to target population, but sampling from community is challenging and resource intense.

Convenience sampling at services and venues (1)

Clients are invited to the study as they attend a service/venue.

- Services aimed specifically at PWID: needle syringe programs, substitution programs, addiction treatment programs other established health services
- Venues: homeless hostels, drop-in centres and social venues/settings

Indicated when services/venues are well accessible for PWID and high coverage

Convenience sampling at services and venues (2)

Advantages

- Low cost
- Logistically simple

Disadvantages:

- Non-random sample: PWID recruited may not be representative for all PWID attending the service/venue
- PWID attending in contact with services/venues can be different in terms of behaviour and prevalence of infectious diseases from the subgroup not in contact with services.
- Especially problematic when venues/services have poor coverage or are not well accepted

Systematic or random sampling at services or from registries (1)

Some services and treatment centres may maintain a registry of users, from which a random or systematic sample can be taken.

Multistage sampling:

- first sampling the services (clusters)
- then sampling the target group members from each selected service unit. The target group members selected from the service unit

Indicated when services/venues are well accessible for PWID and high coverage

Systematic or random sampling at services or from registries (2)

Advantages

- Low cost
- Logistically simple
- Random sample

Disadvantages:

- PWID attending in contact with services/venues can be different in terms of behaviour and prevalence of infectious diseases from the subgroup not in contact with services.
- Especially problematic when venues/services have poor coverage or are not well accepted
- High rate of nonparticipation may result in bias

Convenience outreach sampling

Effort is made to recruit the population possibly not in contact with services (community) through reaching them in open settings

Indicated in case of very stigmatized hidden populations in places where coverage of services is poor and target group members may be reluctant to provide information on their peers.

Advantages:

- May be the first step in getting in contact with a hidden population

Disadvantages:

- non-random sample
- Potentially very biased
- Impossible to estimate bias magnitude and direction
- Reluctance to participate

Time-location sampling (TLS)(1)

Time-Location Sampling is a probabilistic method used to recruit members of a target population at specific times in set venues.

Two-stage sampling approach is used:

- Random sample of time-location units
- Systematic/random selection of participants at each time-location unit.

This sampling method requires formative research and ethnographic mapping to describe key characteristics of the population as well as locations for potential sampling.

Time location sampling (TLS) (2)

Step 1: mapping sites where the target population can be met (includes hours when the population is accessible)

Site 1 and site 2 identified

Step 2: defining time-location units (including exact times, e.g. 4h intervals, 12h intervals)

TLU 1= Site 1 weekday afternoon

TLU 2= Site 2 weekday evening

TLU 3= Site 1 weekend

TLU 4= Site 2 weekday afternoon

TLU 5= Site 1 weekday evening

TLU 6= Site 2 weekend

The fact that the type of person in the location varies by time is not a problem, as the location is included at different times

Time-location sampling (TLS) (3)

Advantages:

- Approaches random sample when assumptions are met

Disadvantages:

- Need of ethnographical assessment (costs ↑)
- bias from non-inclusion of important site(s)
- People who do not congregate in public are most often missed
- Difficulties to interview/collect biological samples in public space:
  - Reluctance to share sensitive information
  - Weather factor
  - Safety issues
- For drug users often the sampling frame overlaps with the drug users in contact with services

Indicated

- when high-risk groups congregate, but their clusters are not stable.
- in case of a relatively open drug scene allowing mapping of places of aggregation of the target population (public venues, open settings)

Targeted sampling

Similar to TLS but sampling frame defined solely by recruitment sites

Sample structure defined by the key characteristics of population described during initial formative/ethnographic phase.

Chain-referal recruitment or direct recrutiment at the site.

Disadvantages:

Potential bias due to fluctuation of population during different hours at the same site

Costly and lenghtly process of ethnographic mapping

Snow-ball sampling

Chain-referral method.

Seeds agreeing to participate in the study are selected. Each of them is asked to provide contact details for other target group members, who can then be contacted by the researcher (random selection of a pre-specified number). This procedure is continued until the desired sample size is reached

Indicated if population is networked and not very stigmatized

Advantages:

- Efficient
- Potential to reach hidden population

Disadvantages:

- Sampling bias from initial seed selection
- Overrepresentation of more cooperative individuals and individuals with larger contact networks

Respondent driven sampling (RDS) (1)

Respondent-driven sampling (RDS) combines "snowball sampling" with a mathematical model that weights the sample to compensate for the fact that the sample was collected in a non-random way.

Known weaknesses of snowball sampling are reduced by

- System of dual incentives (participation bias)
- Recruitment quota and weighting (overrepresentation of PWID with large networks)
- Researchers keep track of who recruited whom and their numbers of social contacts. A mathematical model of the recruitment process then weights the sample to compensate for non-random recruitment patterns

Indicated in case of densely networked populations (average personal network >20 target group members)

Respondent-Driven Sampling (2)

Steps:

- Start with initial contacts or ‘seeds,’ who are surveyed and then become recruiters.
- Each recruiter invites up to three people they know in the high-risk group to be interviewed.
- The new recruits become the recruiters.
- Five to six recruitment waves occur.

Given sufficiently long referral chains (five to six of the people you started with), the final sample will be like the network from which we recruit.

By keeping track of the links between recruiters and recruits and the size of people’s networks, we can calculate the probability of selection and estimate how precisely the population measure is estimated by the sample estimate.

Respondent driven sampling (RDS) (3)

Advantages:

- Efficient
- Potential to reach hidden populations
- Way to correct for non-random sampling

Disadvantages:

- Higher cost
- Bias from non meeting RDS assumptions
- Disconnected subgroups may be missed
- Relatively large samples are needed

Respondent driven sampling (RDS) (4)

There is no scientific consensus whether RDS works or not in practice to produce unbiased estimates.

Even if RDS may not provide unbiased estimates it can be used as an efficient recruitment strategy for obtaining a convenience sample.

- 110 of 123 studies reached at least 90 percent of the intended sample size (on average 263 participants)
- On average, the studies took 9 weeks

Summary (1)

Currently favoured approaches aiming at obtaining an unbiased estimate of the target population:

- Time-Location Sampling

- Respondent-Driven Sampling

But unresolved methodological issues remain.

Service sampling (random or systematic) – allows good probabilty sample from the population in contact with services

How representative is the population in contact with services for the target population?

Convenience sampling (including snow-ball) – useful as first step

Summary (2)

In the framework of surveillance patterns/trends are probably more important than the exact figures/estimates well-funded local research studies may obtain.

Therefore perhaps less rigorous approaches are required

Surveillance demands a systematic, routine approach based on which major indicator trends should be detectable.

For a surveillance system to be adequate and efficient an optimal compromise must be found between representativeness of the data, repeatability of the monitoring system, costs and simplicity.

Summary (3)

In some countries majority of PWID in contact with drug services

- Services access to a sufficiently representative group of PWID
- most surveillance systems will opt for service-based sampling

In countries with low service coverage (< 20-30%) service based sampling may not be appropriate and should be supplemented with community sampling

Discussion

Given the situation in your country what would be the best sampling method for surveillance purposes taken into account representativeness of the data, available resources and feasibility and sustainability of the surveillance system?

Acknowledgements

This presentation is based on the EMCDDA DRID toolkit module ‘EMCDDA (Magdalena Rosinska and Lucas Wiessing). Methods of Bio-Behavioural Surveys on HIV and Viral Hepatitis in People who Inject Drugs — A Short Overview’, EMCDDA, Lisbon, 2013.

This presentation uses slides kindly shared by Prof. Catharina Matheï, KU Leuven, Free Clinic.