Rape Crisis Network Ireland National Statistics Project:

Dr Maureen Lyons,
Equality Studies,
UCD School of Social Justice
Website: www.ucd.ie/esc
Email: Maureen.Lyons@ucd.ie)

"There are lies, damn lies, and statistics. We're looking for someone who can make all three of these work for us."

A brief word about Statistics …

- Statistics: Simply a science of organising and analysing information/data to make the information more easily understood;
- Fictional Example => **Statistical Question** might be:

  How does the service user gender profile of Galway RCC compare to the national picture?

  - Galway Service User Population =500 in 2009, females=400 and males= 100
  - Calculate % female (400/500*100=80%) & % male (%=100/500*100=20%)
  - Nationally the gender spread is 87% Female and 13% Male
  - Statistically (and simply) we say male Service Users are overrepresented in Galway compared to the situation for all RCCs nationally in that year

- First however we must **collect, organise** and **summarise** this information

1. COLLECT & ORGANISE …

- **Collect**: Data from Service Users

- **Organise**: input data into RCNI database, aided by:
  - Dropdown menus & Readily available definitions,
  - Detailed user guide in hard copy.
  - Backup from the RCNI office.

- **Checking, Checking, Checking and Editing!!!** Entry mistakes happen, incorrect information is entered – all of which MUST BE CHECKED and CORRECTED

  Beware: Garbage in , Garbage out …

More about Organising Data

Information entered into RCNI database:

- **Each person using counselling services has 1 record, each helpline contact has 1 record and each accompaniment has 1 record – each record identified by UNIQUE ID**

- **Each survivor using counselling services has the same   number of variables (gender, age, RCC, etc.)**

- **Data Report – selected for time period and available as excel (xls) file**

- **For analysis purposes this is exported into SPSS by the RCNI. This is a specially designed software package for analysing data**

- **Each line is a record where all information relating to 1 person or 1 helpline contact or 1 accompaniment is contained. In all, each survivor counselling record has more than 150 variables**

- **The total no. of counselling records=the total no. of persons utilising counselling, the total no. of helpline records= the total no. of helpline contacts, the total no. of accompaniment records=the total no. of accompaniments.**

Characterising RCNI Data

- RCNI information classified as Administrative Data
- This mean that Data originate from the 'activities/daily work' of an organisation' i.e., in the context of face-to-face counselling, accompaniment and the helpline
- Definitions used are driven by administrative and not research purposes – this has positives and negatives
- Designated person(s) within organisation is responsible for all aspects of data production

Advantages of Administrative data

- **No further burden** on Service Users
- **Low cost** compared to other methodologies **e.g. regular censuses/surveys**
- **Complete coverage** [census] of population of Service Users
- **Good possibilities for reporting** for individual RCCs, small areas, regional statistics as well as national statistics
- Additional **long term possibility of longitudinal data**

& any Drawbacks?

- Restrictions on data collected:
  - Population – data only representative of the client/service user group of the organisation itself
  - Information collected – questions/portions of individual story that are appropriate to Helpline or Counseling or Accompaniment;
- Variable definitions can change as developments/ improvements happen - result is that these data lose comparability over time period
- Variables that are less important for administrative work can be of lower quality yet these can often represent the most important for statistical reporting purposes
- And again Garbage in … Garbage out!

- A lot of the variables (such as gender, age, etc.) are descriptive and so straightforward and unproblematic once the information is correctly coded.

- It is only by cross checking that it is possible to spot data entry mistakes

  - e.g., Gender of survivor recorded as male and also positive recording of pregnancy following rape.

  - Obvious difficulty which can be seen for all Service Users by running a crosstab of Pregnancy By Gender, where all occurrences for male should be zero

  - A comprehensive list of such data cleaning/editing can be produced and run each year when preparing data for export.

  - This can also be done manually for each person but is more time consuming and also more prone to error

3 issues in the process of data development

1. What questions can we address given the specific context? Consider for e.g.: the actual population and what this means in terms of generalising from the data; the implications of time lapse since abuse; the difficulty of capturing the complexity of aspects of the abuse

2. What precise data do we need to answer these questions? e.g.,To provide data on the 'mean number of perpetrators we need to specifically identity each perpetrator with a unique id

3. What specific analysis/analyses do we need to conduct in order to address our questions. e.g. if we are just interested in responses from female service users then we must segment the data to include only women and calculate %s for this group; do we need to include or exclude missing cases?

Example of how and why changing data is part of data process

- Nature of abuse in terms of a) age of survivor when each incident of abuse began and stopped; 2) how many abusers were involved for each incident and 3) relationship between survivor and abuser

- Just to describe the nature of abuse experienced by a survivor we need an elaborately constructed sequence of data which is correctly recorded in terms of distinct incidents – this requires information on age began, age ended, gender and age of each uniquely identified perpetrator as well as the relationship between each survivor and perpetrator.

- When the database was constructed we did not realise the importance of
  - uniquely identifying each perpetrator; and
  - Recording the precise age abuse began and ended in each particular abuse situation

  Using age bands for the age abuse began and not identifying each perpetrator uniquely meant we were unable to describe the nature of the abuse incident in the way required for these data to be accurate and meaningful

How we overcome shortcomings of initial 'nature of abuse' data

- Survivor – ask specific age each incident of abuse began and ended
  - This allows us to accurately calculate how many incidents for each survivor
  - this was not possible previously when age bands were used which resulted in multiplication of incidents if abuse spanned more than one age band
- Perpetrator – now using ' perpetrator id' variable –
  - this allows us to accurately calculate no. of perpetrators and relationship between survivor and perpetrator unlike before when we could not say with an accuracy whether, for example, the uncle recorded in <u>abuse incidents 1 and 3</u> referred to 1 or 2 uncles.
  - This has now been rectified and by using a particular command syntax in SPSS we can, in the future, calculate the accurate total counts of perpetrators per incident and overall for each survivor.

Final Word on Reporting

- ## Reflect what data actually capture
  - RCC Service Users: Sub-group of total in current population who 'ever experienced Sexual Violence'
  - Year on year analysis provides a Snapshot view of Service Users of RCCs in that calendar year, it does NOT reflect incidence of sexual violence in that year;
  - Cannot make general statements about sexual violence – for example, comment on how nature of sexual violence has changed in society or whether incidence has increased/decreased