CanReg analysis improvements

Morten Johannes ERVIK, Freddie BRAY

April 26, 2011

1 Introduction

R is a powerfull open source freely available software package that could be coupled with CanReg to improve the analytical capabilites.

2 What data does registries store in CanReg5

Each record store (at least):

- Sex
- Incidence date
- Birth date (or age at the time of tumour)
- Coded address at the time of the tumour
- Topography, Morphology, Behaviour in ICD-O-3
- Most valid basis of diagnosis
- ICD10 and ICCC
- Date of last contact
- Vital status
- Source info:
  - Type of
  - Number of
  - (Dates)

Globally each registry also have population data sets (denominators).

3 What we *do* with that data now

- Incidence tables (Per 100.000 per cancer group, age group, ASR, CR etc.)
- Number of cases (per cancer group, age group)
- Population pyramids
- Frequencies by year

Otherwise data needs to be exported to be analysed in other software packages.

4 What we *could* do with that data

- Graphs:
  - Bar charts by cancer/sex
    - incidence tables
    - number of cases
  - Time trends
    - ASRs (world) over time
    - age spec rates over time
    - age spec rates over cohort
- Quality indicators
  - Validity:
    - DCO%
    - PSU%
    - MV%
    - Compared with other reg (CI5 IX)
    - DCO% over time
      - potentially with graphs
  - Completeness:
    - Reference childhood incidence comparison
    - Stability of rates over time by cancer/sex
      - potentially with graphs
    - Age specific rates by cancer/sex
    - Sources
      - number of sources per case
      - number of notifications per case
- Geographic stuff, maps?

5 What we could do if we linked it to other data

This is more for the future, but might be interesting...

- If linked to mortality data
  - M/I ratios as estimator of completeness

6 Technical aspects

Basically two main ways to do it.

6.1 Export the data from CanReg5 to files readable by R and then call R in batch mode

Prefered mehtod - more dynamic and loosely coupled. Easier to potentially reuse R code later.

6.2 Use libraries in Java (a Java to R bridge) to call R functions directly

Alternative method we might want to look into.