CanReg analysis improvements

Morten Johannes ERVIK

April 13, 2011

1 What we have

Each record store (at least):

- Sex
- Incidence date
- Birth date (or age at the time of tumour)
- 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
- Sources
  - Type of
  - Number of

Globally we also have population data sets (denominators).

2 What we *do* with that data now

- Incidence tables (Per 100.000 per age group, ASR etc.)
- Number of cases
- Population pyramids

3 What we *could* do with that data

- Graphs
  - Time trends?
  - Cohort-thingies?
- Survival analysis?
- Quality indicator graphics? By year/period/user/anything?
  - Number of sources per record per type?
  - Microscopically confirmed ratios?
  - DCO% per type?
- Geographic stuff, maps?

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

- Mortality data
  - Quality indicators
  - Completeness
  - M/I ratios

5 Technical aspects

Basically two main ways to do it.

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

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