Statistical analyses of the German uranium miner cohort

Christina Sobotzki

Federal Office for Radiation Protection /

Bundesamt für Strahlenschutz (BfS)

Neuherberg, Germany, E-Mail: csobotzki@bfs.de

DoReMi LD-RadStats Workshop 2015, Barcelona

German uranium miner cohort study

- **Study population:**
  N = 58,982 former miners of the uranium mining company "Wismut"
- **Follow-up period: 1946 – 2008**
  Duration of follow-up: Ø 37 years
  Person-years at risk: 2,180,700
- **Number of deaths:**
  Total: n = 25,438
  Lung cancer: n = 3,500
  Cancers other than lung cancer: n = 4,278

Wismut underground miners, 1950ies

Data basis

Dataset with 58,982 rows (one row per person)

Basic-Information

- ID
- Date of birth
- Begin and end of follow-up
- Vital status according to date of end of follow-up
- Cause of death (ICD-10-Code)

Information of the employment

- Start of employment
- End of employment
- Workplace (surface/underground/milling/open pit)

Information on the exposure

Annual values for 1946-1989:

- Radon and its progeny [WLM]
- External gamma radiation [mSv]
- Long-lived radionuclides [kBq h/m^3]
- Silica dust [mg m^{-3}-Jahre]

(WLM: cumulative exposure of alpha emitters from radon progeny)

Objectives

Main objectives:

Based on current follow-up period (1946-2008) the main objective of the analyses of the German uranium miner cohort study is to investigate associations between mortality from lung cancer, cancers other than lung cancer, non-malignant respiratory diseases and cardiovascular diseases and occupational radiation exposure to:

- Radon and its progeny
- External gamma radiation
- Long-lived radionuclides
- Silica dust

Objectives of this talk

What is the appropriate method for statistical analysis?

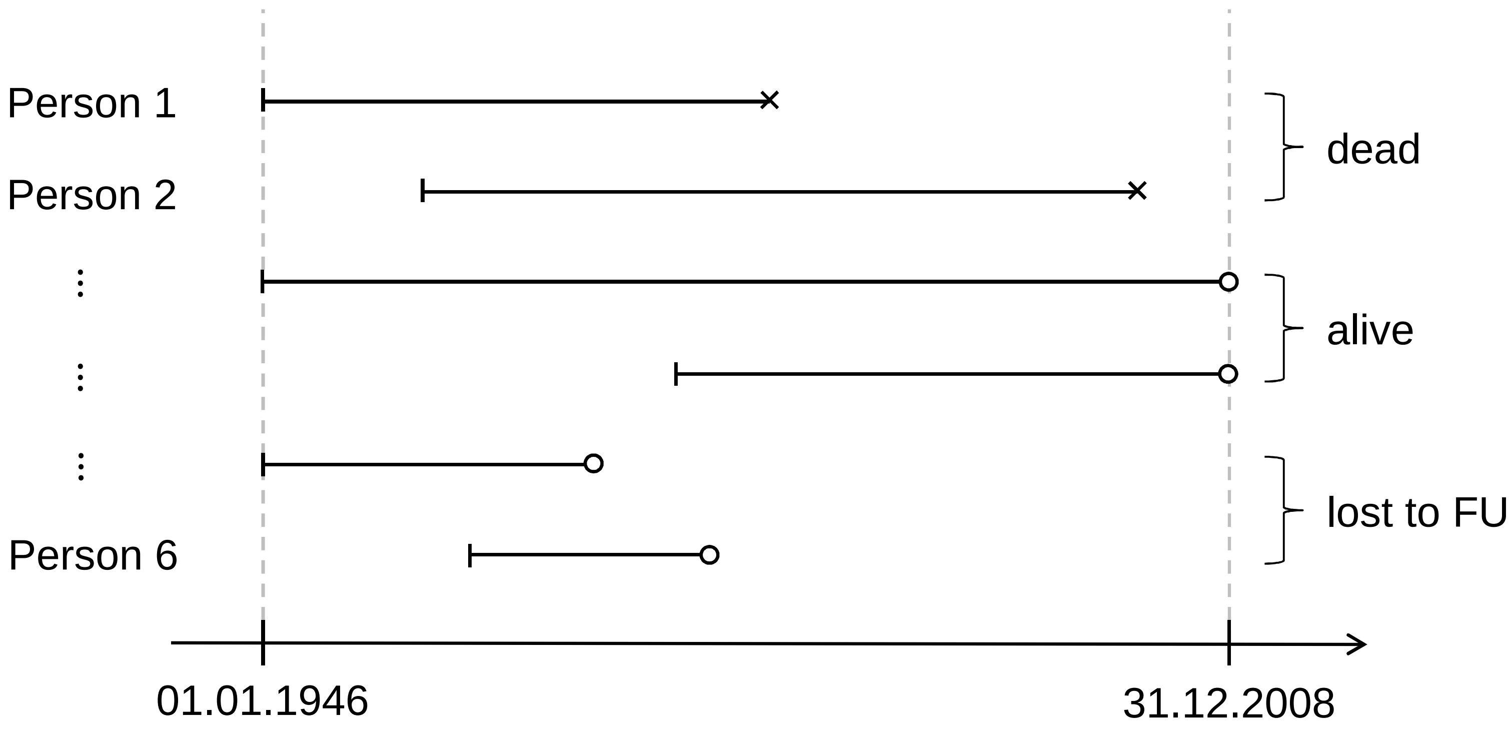
- Internal Poisson regression as established method

Requirements for this method:

- Estimation of the assocation between our outcome (mortality) and the exposure of interest
- Time components must be considered
- Complex data structure of our cohort
- Statistical Software

Do we have any alternatives?

Data structure of the Wismut cohort



„Trick“ to get count data: grouped data set

	Year *y*						
Age *a*	1946	1947	1948	…	2006	2007	2008
0-15							
15-20							
20-25							
⋮							
⋮							
75-80							
80-85							
>85							

Calculation per cell:

N Number of persons

pyr Number of person-years

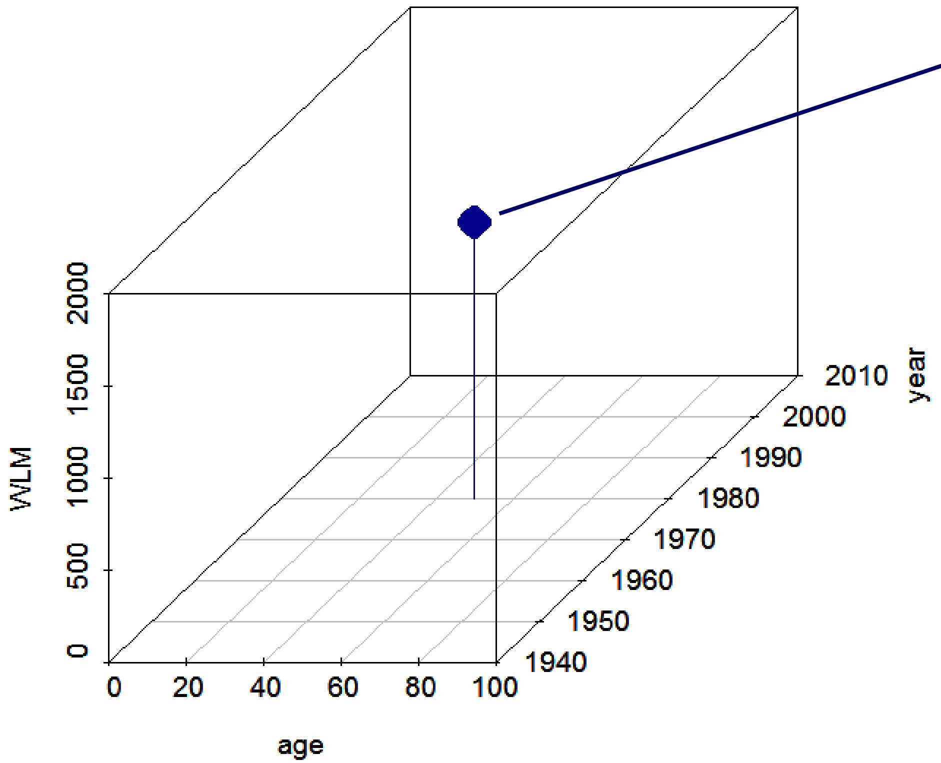
n Number of deaths

Baseline mortality rate

per age group and year:

$$r_0(a, y) = \frac{n}{pyr}$$

Grouped data set – considerung radiation exposure



Calculation per cell:

N Number of persons

pyr Number of person-years

n Number of deaths

Cancer mortality rate

per age group, year & radiation exposure**:**

$$r(a, y, rn) = \frac{n}{pyr}$$

Classical Statistical Analysis

- Internal Poisson regression
- Estimation of Excess Relative Risk (ERR)

ERR model for radon *rn* :

$$r(a, y, rn) = r_0(a, y) \cdot [\ 1 + \underbrace{ERR(rn)}_{\beta \cdot rn}\ ]$$

- $r(a, \mathrm{y}, rn)$ = cancer mortality rate at year (y) and age (a), after a radiation dose (e.g. rn)
- $r_0(a, y)$ = the age and year specific baseline mortality rate

$$\leftrightarrow \quad \frac{r(a,y,rn)}{r_0(a,y)} = 1 + ERR(rn) = \mathrm{RR}$$

Statistical Analysis

Simple Model:

- Categorical (grouped cumulative exposure): (0-10, 10-50, 50-100, 100-200, 200-500, 500-1000, 1000-1500, >1500 WLM)
- Linear (ERR per unit of cumulative exposure)
- Stratified for age (16 categories) and calendar year (5-year groups)
- Lagged exposure to consider latency
    - ➢ Calculation of the ERR per lagged cumulative exposure units

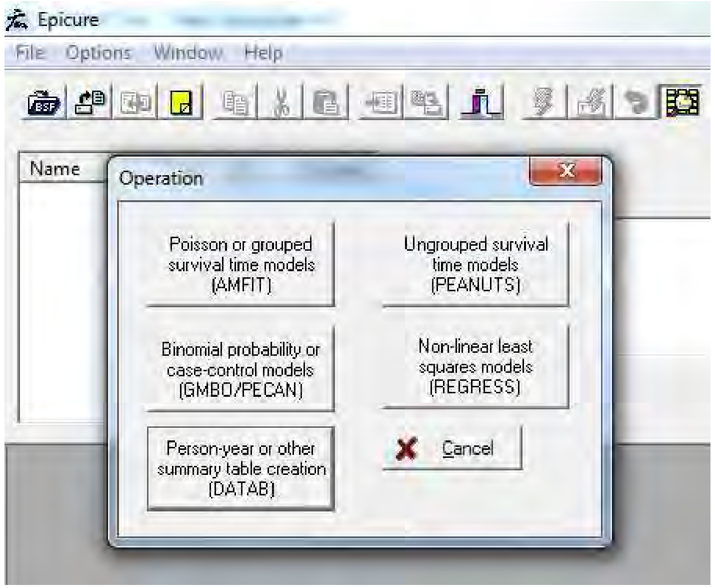
Advanced Model:

- Further consideration of effect modifiers:
    - ➢ Age at exposure, Time since exposure, exposure rate

Software

Epicure

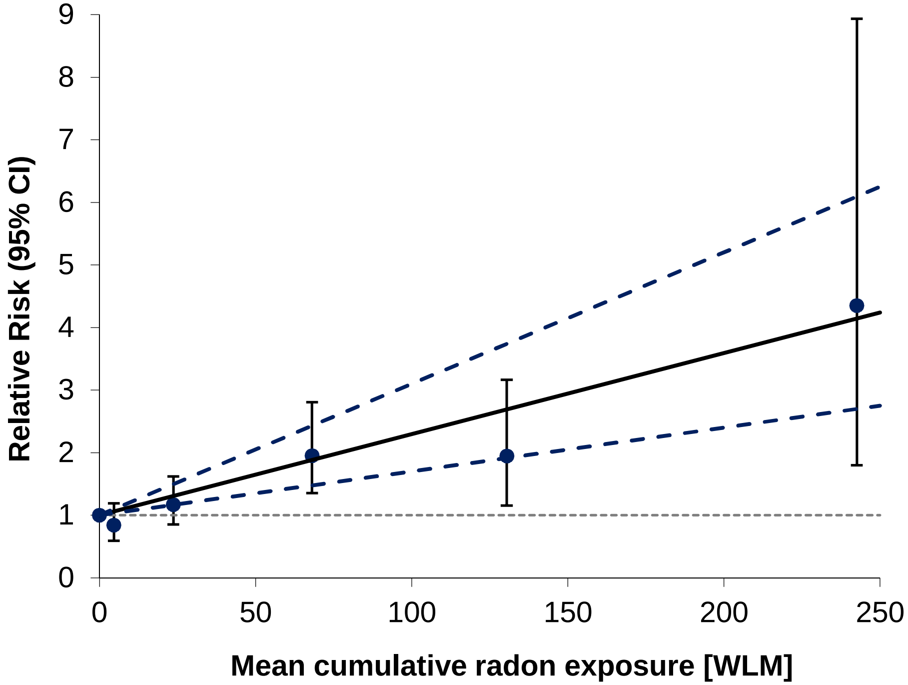
- Very specialised software with own programing language
- Rather old (1990th)
- No maintenance / support
- No community (e.g. forums)
- Fast and well working program




```
! Enlarge workspace size
new size 8000000 @

LOG 0_datamanagement.log @

!-----------------------------------------------------
! Variable preprocessing
!-----------------------------------------------------

! Change yyin and yyout as needed

TRAN IF yystop > 1989 THEN
yystop=1989;
mmstop=12;
ddstop=31;
ENDIF;
@

TRAN IF yyin < 1946 THEN
yyin=1946;
mmin=1;
ddin=1;
ENDIF;
@

! Set lag time for leukaemia

CONSTANT #lag=2 @


!-----------------------------------------------------
! Initialize arrays for cumulative lagged doses
!-----------------------------------------------------

ARRAY \highLETdosis   dhlt1946-dhlt2008 @
ARRAY \lowLETdosis   dllt1946-dllt2008 @
```

Results

Lung cancer risk from cumulative radon exposure in the sub-cohort of miners hired after 1960 (n=26.766)



n = 334 (lung cancer deaths)

ERR/WLM = 0.012 (95% CI: 0.007; 0.021)

Statistical significant relationship between death from lung cancer and cumulative radon exposure

↯ Relative Risk ≙ (Mortality) Rate Ratio

Strengths & Limitations

Strength:

- Established method

Limitations:

- Complex model
  - ➢ Difficult interpretation
- Data must be grouped in the preliminary stage
  - ➢ Loss of information
- Slightly different risk estimates depending on grouping

Grouping Variable	β (95% CI)
Low LET (0,]0-3.69],]3.69-12.8],]12.-41.0],]41.0-197],>197)	0.37 (-1.63;2.37)
Low LET (0,]0-10],]10-25],]25-50],]50-100],] 100-250],]250-1000], >1000)	0.25 (-1.65;2.15)
Low LET (0,]0-3.69],]3.69-12.8],]12.-41.0],]41.0-197],>197) + radon, gamma & LRN	0.58 (-1.42;2.59)
Low LET (0,]0-10],]10-25],]25-50],]50-100],] 100-250],]250-1000], >1000) + radon, gamma & LRN	0.31 (-1.58;2.20)

Statistical Analysis

Idea: Use survival analysis methods for individual data

- We have a common survival data structure
- Method can estimate the association between radiation exposure and mortality risk in continuous time
- Can account for censored and stratified data

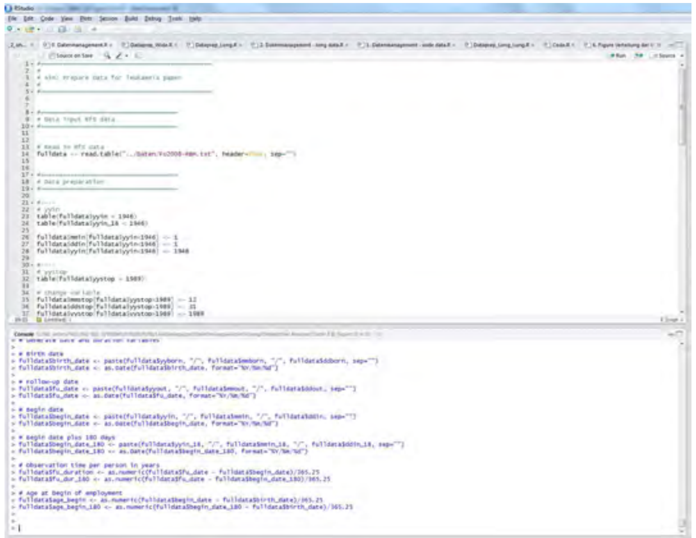
e.g. Cox model for radon *rn* :

$$\lambda(t, rn) = \lambda_0(t) \cdot \underbrace{\exp[\beta \cdot rn]}_{Relative\ Risk} \quad \leftrightarrow \quad \frac{\lambda(t, rn)}{\lambda_0(t)} = \exp[\beta \cdot rn]$$

- $\lambda(t, rn)$ = hazard function of time (t) and covariates (e.g. rn)
- $\lambda_0(t)$ = the baseline hazard function

But: we do not estimate the ERR

Suggestion: develop R-package and implement ERR approach in cox regression models



Strength:

- Free statistical software
- Good support, large community (e.g. forums)
- No grouped data necessary
  - ➢ No loss of information
  - ➢ Could lead to smaller confidence intervals
- Estimate time-depedent effect modifiers in individual data
- Possibility to model non-linearity with e.g. p-splines

Limitations:

- R: programming language
- Complex: own preparation of dataset to longitudinal format

Summary

- Internal Poisson regression estimating ERR and using Epicure is established in the radiation research
- But grouping the individual data is necessary accepting a loss of information
- Cox regression with individual data could be an alternative method
- R als statistical software seems promising for future analyses

Thank you for your attention!