5.3 Weighted Cox regression of nested case-control data

By ”Breaking the time-matching”, we mean

- Reweight the nested case-control individuals
- Thus reconstruct the person-time structure (number of individuals at risk at different times) of the whole cohort
- Use weighted Cox regression
- Valid estimates of (incl. for matching factors)
- Can estimate absolute risk

Advantages of weighted Cox regression vs. conditional logistic regression)

- Overcomes loss of concordant sets
- The (weighted) controls can be used as a comparison group for another outcome/disease of interest in the same cohort
- Can estimate HR for the matching factors
- Enables estimation of the absolute risk

Study of postpartum VTE

rare exposure(s): transfusion

	Cohort	1:5 NCC CLR	1:5 NCC IPW
	N=966,070		
	472 cases	81 discordant sets	all 472 sets
RBC units:			
1-2	2.53(1.57,4.07)	2.69(1.45,4.98)	2.58(1.61,4.15)
3-5	2.79(1.44,5.42)	3.06(1.17,8.03)	2.42(1.20,4.85)
¿5	4.36(1.62,11.7)	3.65(0.87,15.3)	4.03(1.49,10.9)
Smoking	1.51(1.13,2.03)	1.42(1.01,2.01)	1.42(1.06,1.92)
Preeclampsia	2.50(1.79,3.48)	2.15(1.37,3.36)	2.13(1.50,3.04)
Delivery:			
Instrumental	1.08(0.76,1.54)	1.18(0.80,1.76)	1.16(0.81,1.66)
Elective CS	1.73(1.28,2.36)	1.76(1.23,2.52)	1.74(1.28,2.37)
Emergency CS	2.26(1.77,2.89)	2.41(1.78,3.27)	2.37(1.84,3.05)

Weighted analysis allows:
reuse nested CC data for a new outcome

The reweighted data represents the full cohort

So we can use it to do any analysis that we could do with the full cohort.

Example of reusing controls: The contralateral breast cancer (CBC) study

Background

- Contralateral breast cancer definition:
  second primary breast cancer in the contralateral side,
  detected at least three months after the first breast malignancy

- Risk factors for CBC:
  Known risk factors:- family history, - non-ductal histological type
  - young age at diagnosis of the initial breast cancer

- Investigated but often reported as non-significant:- parity

- Never investigated:
  - multifocality of the initial breast cancer (BC) tumor

Contralateral breast cancer (CBC)

Research question

Is multi-focality of the first breast cancer a risk factor for CBC?
Is parity a protective factor for CBC?

Data

- Patient cases of CBC identified in Stockholm-Gotland Cancer Register (1976-2005).
- Variables of interest retrieved from medical charts
- 853 CBC cases

Example: Contralateral breast cancer (CBC)

853 CBC cases

- Could collect controls in a NCC design
- Time and cost
- reuse control data from another NCC study
  - → cases of metastases subsequent to BC (1997-2005)
  - → controls sampled in a NCC design
  - → matched on intended treatment, age category (<45, 45-54, >54 years), and binary variable (treatment < 2001 or ≥ 2001)

- Same variables retrieved from medical charts as for CBC cases

Both studies within the same cohort

Cohort: All breast cancer patients 1976-2008

Metastases study:

BC between 1997-2005
< 76 years old
Treatment include chemo- or hormonal therapy
matched NCC

CBC cases:

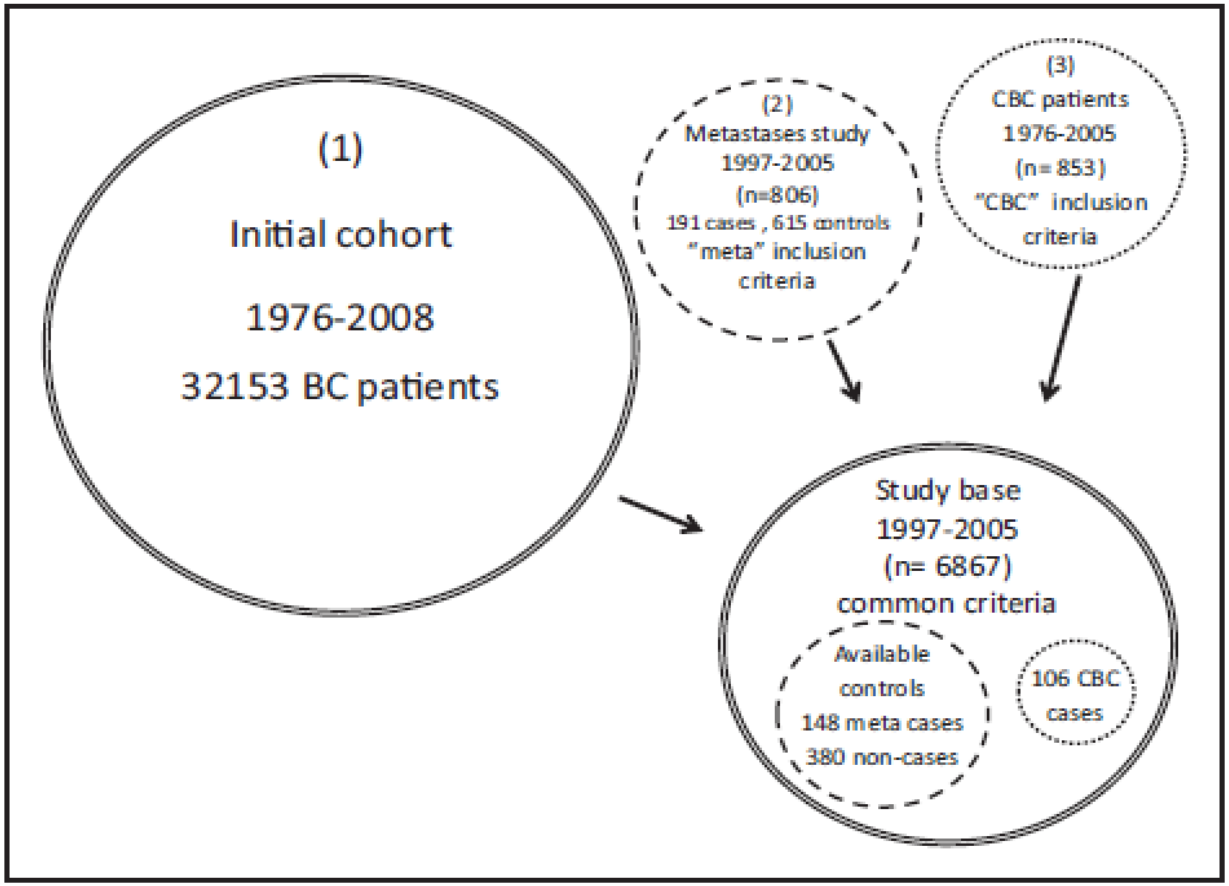
BC between 1976-2005
> 3 months between BC and CBC
No prior malignancy

Different inclusion criteria! (so not the same ”study base”)

How to correctly use the data?

- We need to identify a common “study base”
- can be reconstructed from the Stockholm Breast Cancer Register which recorded a total of 32 153 BC patients from 1976 to 2008

“Aligning” the data sets



Results

Table 2. Adjusted risk estimates: hazard ratios (HR) and 95% confidence intervals (CI) from Cox regression analyses.

Risk factors	Main analysis[a]	Sampled controls[b]	Study base for metastasis[c]	Unweighted[d]
Non-multifocal tumor (ref.)	1	1	1	1
Multifocal tumor	1.99 (1.07, 3.70)	2.03 (1.08, 3.81)	1.98 (1.07, 3.70)	1.60 (1.06, 2.41)
Nulliparous (reference)	1	1	1	1
Parity	0.40 (0.18, 0.89)	0.37 (0.16, 0.85)	0.39 (0.18, 0.86)	0.79 (0.47, 1.33)

details in:

Article

SMMR STATISTICAL METHODS IN MEDICAL RESEARCH

Statistical Methods in Medical Research
0(0) 1–12

DOI: 10.1177/0962280216669744
smm.sagepub.com
SAGE

Feasibility of reusing time-matched controls in an overlapping cohort

Bénédicte Delcoigne,[1] Niels Hagenbuch,[1] Maria EC Schelin,[2] Agus Salim,[3] Linda S Lindström,[1,4] Jonas Bergh,[1] Kamila Czene[1] and Marie Reilly[1]

Results

Table 2. Adjusted risk estimates: hazard ratios (HR) and 95% confidence intervals (CI) from Cox regression analyses.

Risk factors	Main analysis[a]	Sampled controls[b]	Study base for metastasis[c]	Unweighted[d]
Non-multifocal tumor (ref.)	1	1	1	1
Multifocal tumor	1.99 (1.07, 3.70)	2.03 (1.08, 3.81)	1.98 (1.07, 3.70)	1.60 (1.06, 2.41)
Nulliparous (reference)	1	1	1	1
Parity	0.40 (0.18, 0.89)	0.37 (0.16, 0.85)	0.39 (0.18, 0.86)	0.79 (0.47, 1.33)

Does it matter if use the controls only from the metastases study or all data?

No, it is not a problem.

Weights aim to reconstruct the cohort.

Among the controls there were metastases cases, who were weighted => should represent similar patients in the cohort.

Cases only get weight 1, much less influence than controls

If ”reconstruction” of the study base is imperfect

Not a problem when the study base is large.
Risk sets are large even at the end => not much influence on the product in KM type weights.
Could be problematic for small study bases with stratification.

Does it matter if?

We use unstratified weights instead of stratified ones?

It should not be a problem as the main factor that influences the weights is the time.

Adjustment for matching factors may compensate for using unstratified weights, but effect of those factors likely biased?

Does it matter if?

We use date of selection instead of last date of follow-up?

Highly problematic: influences both weights and likelihood

Considerations for re-using biomarker data

Potential confounding with storage time and technology:

Storage Time:

- Single entry time, nested CC study already done, now new cases of interest, conduct small nested CC study and supplement?

- Staggered entry, storage times may overlap, model?

Technology: calibrate old controls with validation sample

Weighted analysis of nested CC data: allows estimation of absolute risk

Absolute risk estimation for cohort data:

Breslow estimator *

$$H_0(\mathrm{t}) = \sum_{i=1} \frac{I\,(t_i \le t)}{\sum_{k \in R_i} \exp[\beta X_k + \gamma Z_k]}$$

For nested case-control data: adapted Breslow estimator

$$H_0(\mathrm{t}) = \sum_{i=1} \frac{I\,(t_i \le t)}{\sum_{k \in R_i^{\#}} \exp[\beta X_k + \gamma Z_k]\, w_k}$$

w_k Kaplan-Meier type weight

* *Hanley J, Epidemiology 2008*

Application in cancer research:

risk of developing lung cancer after radiation treatment for breast cancer

- Radiation therapy may increase risk of lung cancer
- Particularly in smokers
- Interaction between smoking and radiotherapy??

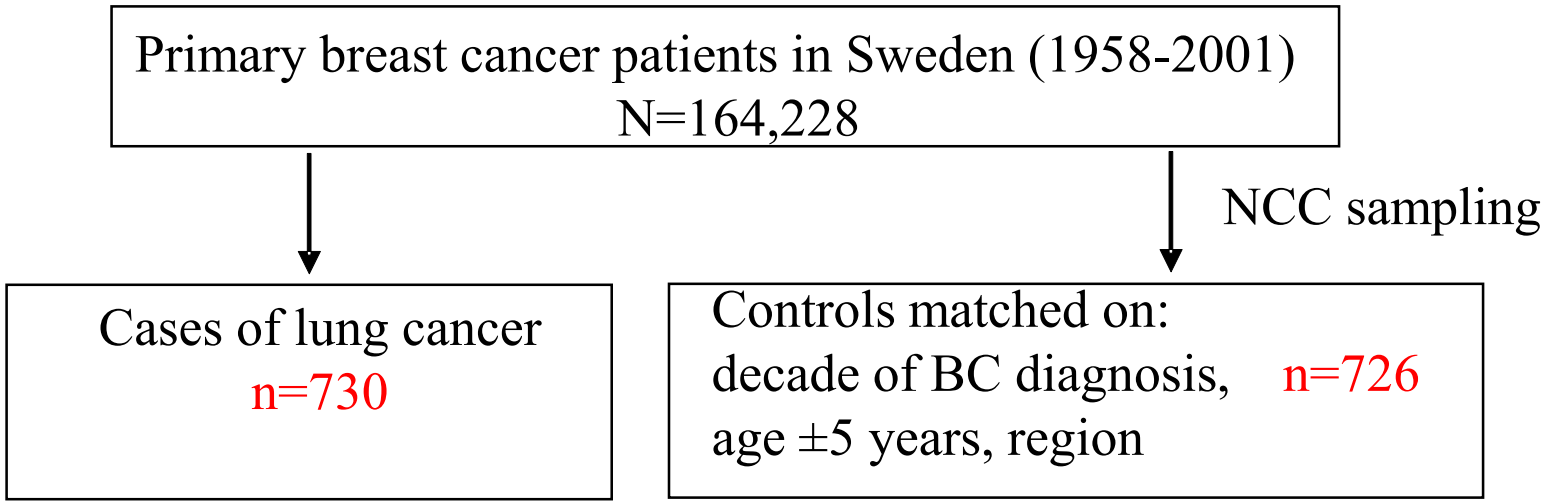
More on data available

- Dates & clinical details for breast (and lung) cancers,
- Laterality (L/R) of the cancer(s)

- Radiation doses received at each lung

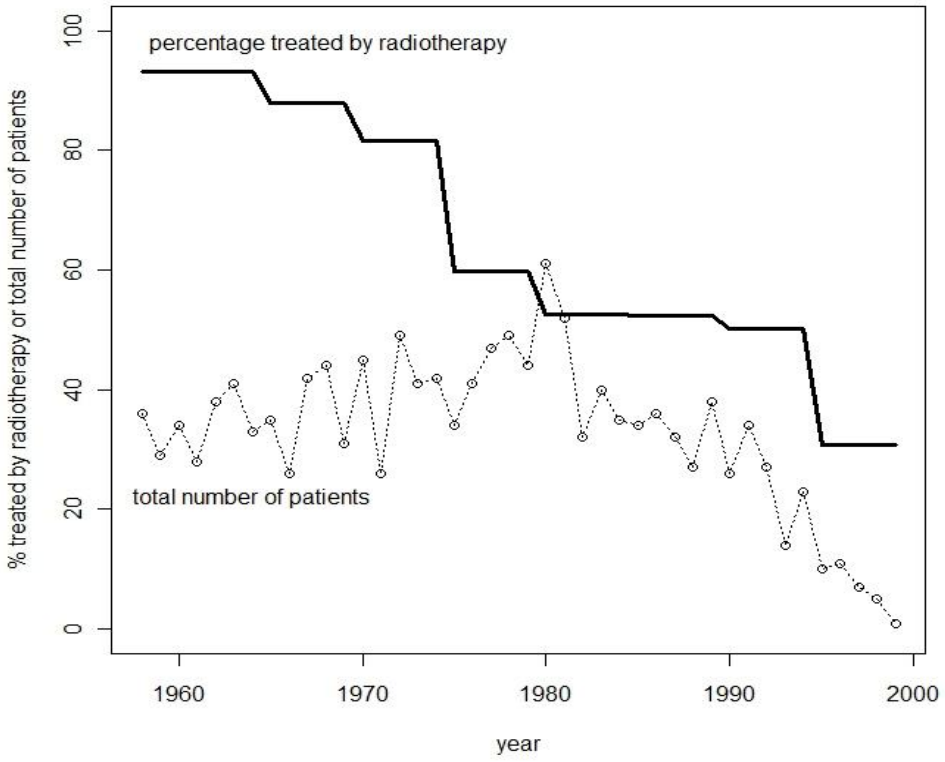
- Smoking information

Design & data



Conditional logistic regression – no significant effect of radiation

Radiotherapy use by calendar year: overmatched?

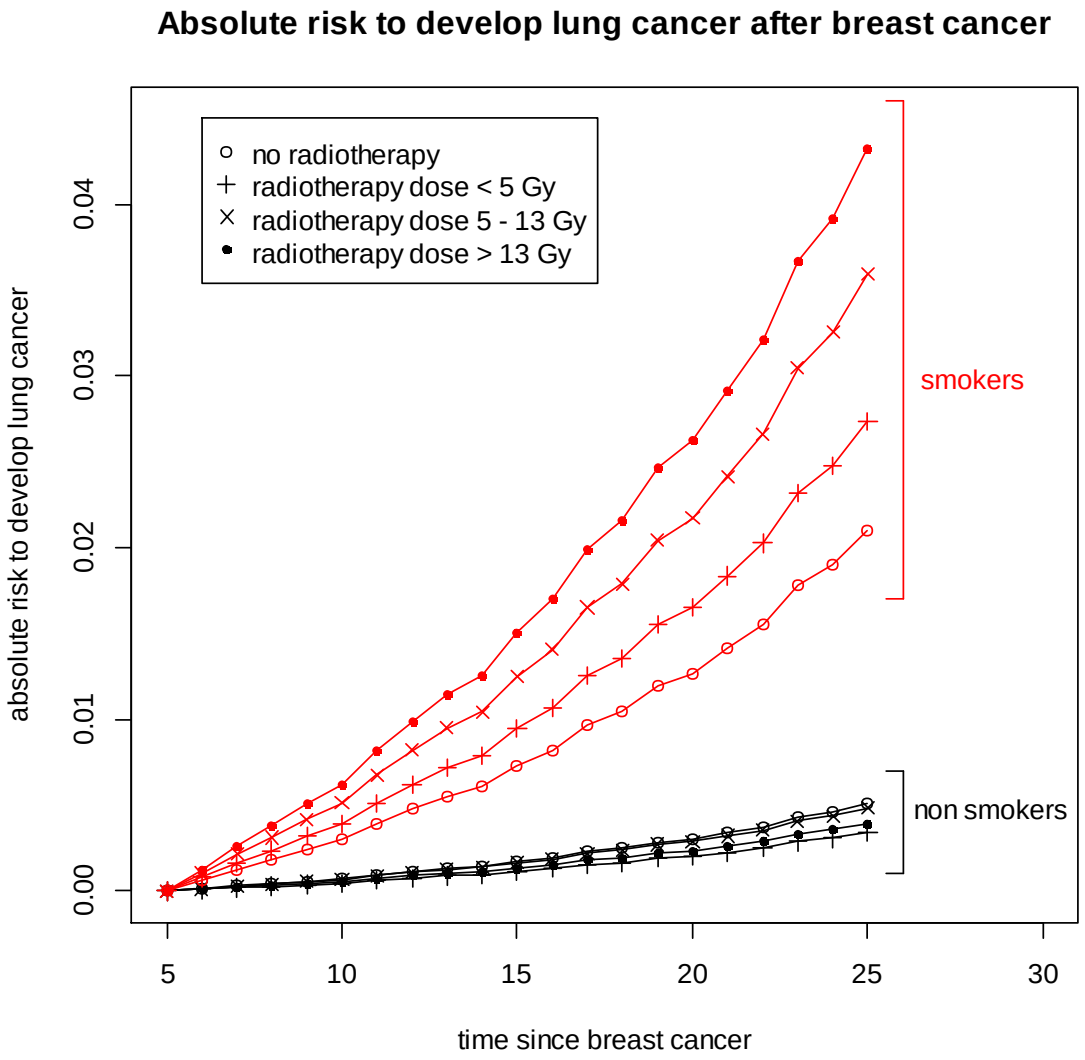


Breaking the matching

- Solve problem of overmatching on calendar time
- Allow use of data on individual lungs!
- Calculate absolute risk

Results (adapted Breslow estimator)

absolute risk of Lung Cancer
- for a 54 year old woman



Conclusion:

Radiotherapy is a risk factor

Effect is modified by smoking

Evidence of a dose-response effect of radiation dose in smokers

Many potential advantages from breaking the matching in nested case-control data

Exercise 5.2