Approaches to handle missing follow-up time: A comparative analysis of contralateral breast cancer incidence

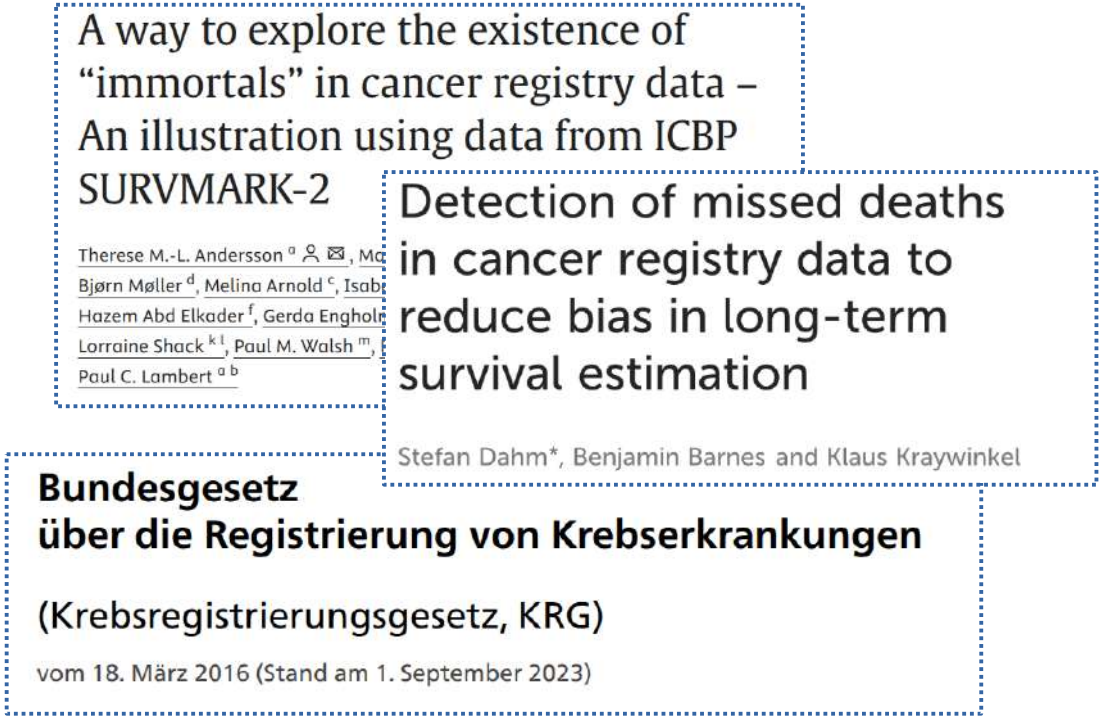
Sarah Haile, Miriam Wanner, Dimitri Korol, Sabine Rohrmann

Epidemiology, Biostatistics and Prevention Institute (EBPI),
University of Zurich, Zurich, Switzerland;
Cancer Registry of the Cantons of Zurich, Zug,
Schaffhausen and Schwyz, Switzerland

GRELL & ENCR – 2025 Scientific Meeting

The problem: missing vital status or follow-up time

- Complete vital status data is essential in cancer registries
- Not always possible due to missing or incorrect linkage
- “missed deaths” / “immortal” subjects stick around in the cancer registry, even if the situation has improved
- How do we handle these cases properly?



Potential solutions

- Ignore it
- exclude
- “simulation” / Single imputation
- Multiple imputation
- ...

and 2015. We excluded women with a prior cancer, women diagnosed at age <30 or age 8
and older, had unknown laterality, undocumented nodal status, invasive cancers with
histology not of ductal, lobular or mixed (ductal–lobular) subtype, in situ cancers with a
histology not of ductal, discordant data entry in the SEER database (i.e. T stage classified a
at least T1, but stage was classified as stage 0), missing follow-up, had bilateral breast
cancer or stage IV disease. We excluded women who had no surgery and those who had a

i.e. second non-breast cancer, death or end of study period (31 December 2011). Because of insufficient follow-up in most cases, person-years was simulated with Swiss breast cancer survival data provided by NICER. A constant hazard rate of 0.0575 y^{-1} for survival after UBC was assumed for the whole period from 1980 to 2011 to compute random survival time.

Multiple imputation			Simulated 20× using based on the following variables:
	A	MI-A	Age, age^2, period, period^2, language region (German, French, Italian)
	B	MI-B	Age, age^2, period, period^2, language region (German, French, Italian), vital status (δ), baseline hazard, presence of multiple primary tumors
	C	MI-C	Age, age^2, period, period^2, language region (German, French, Italian), presence of multiple primary tumors
	D	MI-D	Age, age^2, period, period^2, language region (German, French, Italian), vital status (δ), baseline hazard, presence of multiple primary tumors

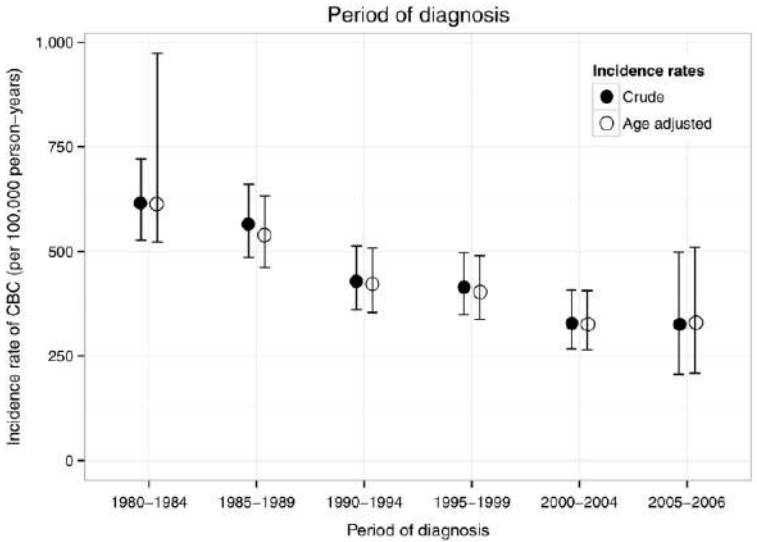
Better tools for better estimates: improving approaches to handling missing data in Swiss cancer registries

Cornelia Richter[a,b,c], Lea Wildisen[b,c], Sabine Rohrmann[a,b] and Sarah R. Haile[a]

Application to incidence of contralateral breast cancer:
update previous analysis - 10 years later

Incidence of metachronous contralateral breast cancer in the Canton of Zurich: a population-based study of the cancer registry

Julia Prater[1] · Fabio Valeri[1,2] · Dimitri Korol[1] · Sabine Rohrmann[3] · Silvia Dehler[1]



Cantons of Zurich and Zug:

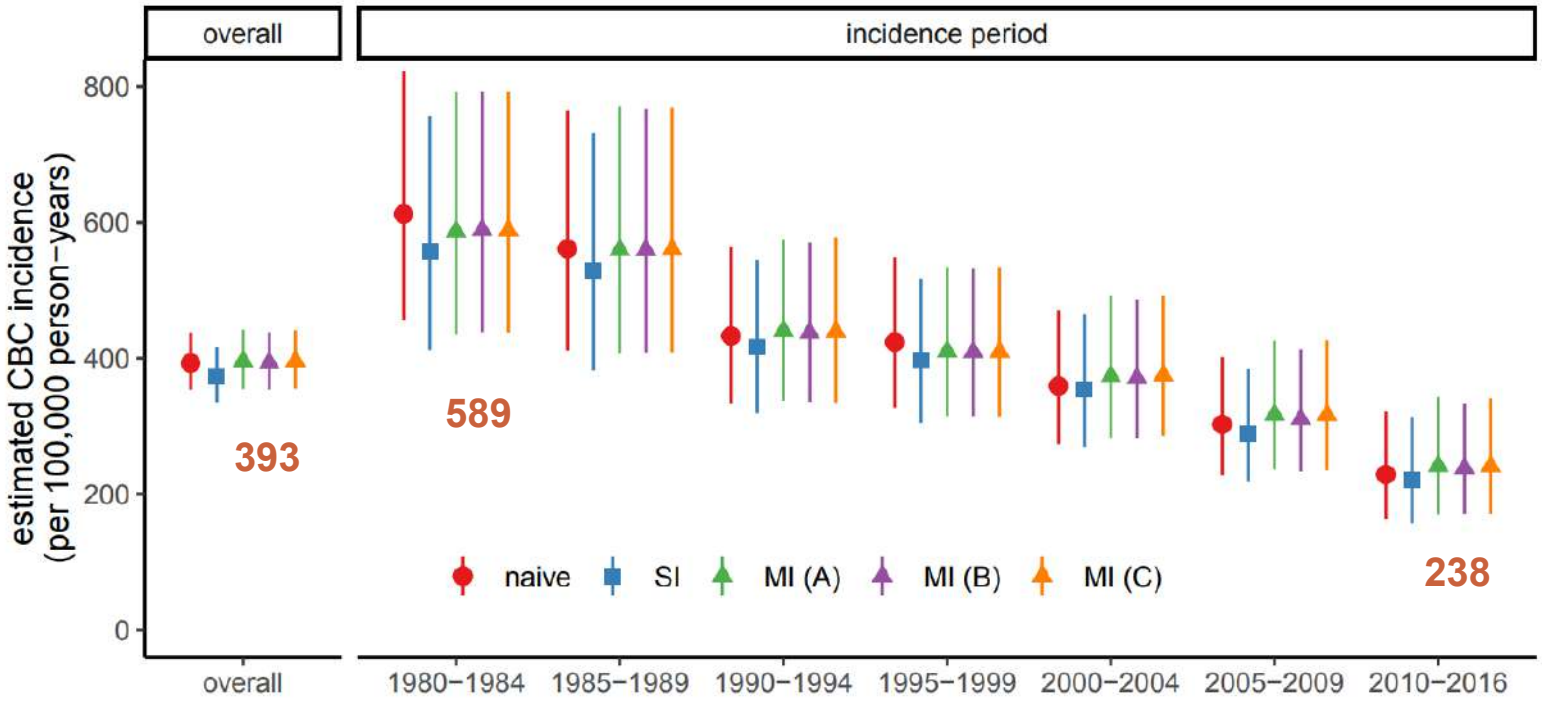
- 1.7M inhabitants,
- 20% of Swiss population,
- cancer registry since 1980 (Zug 2011)

Methods

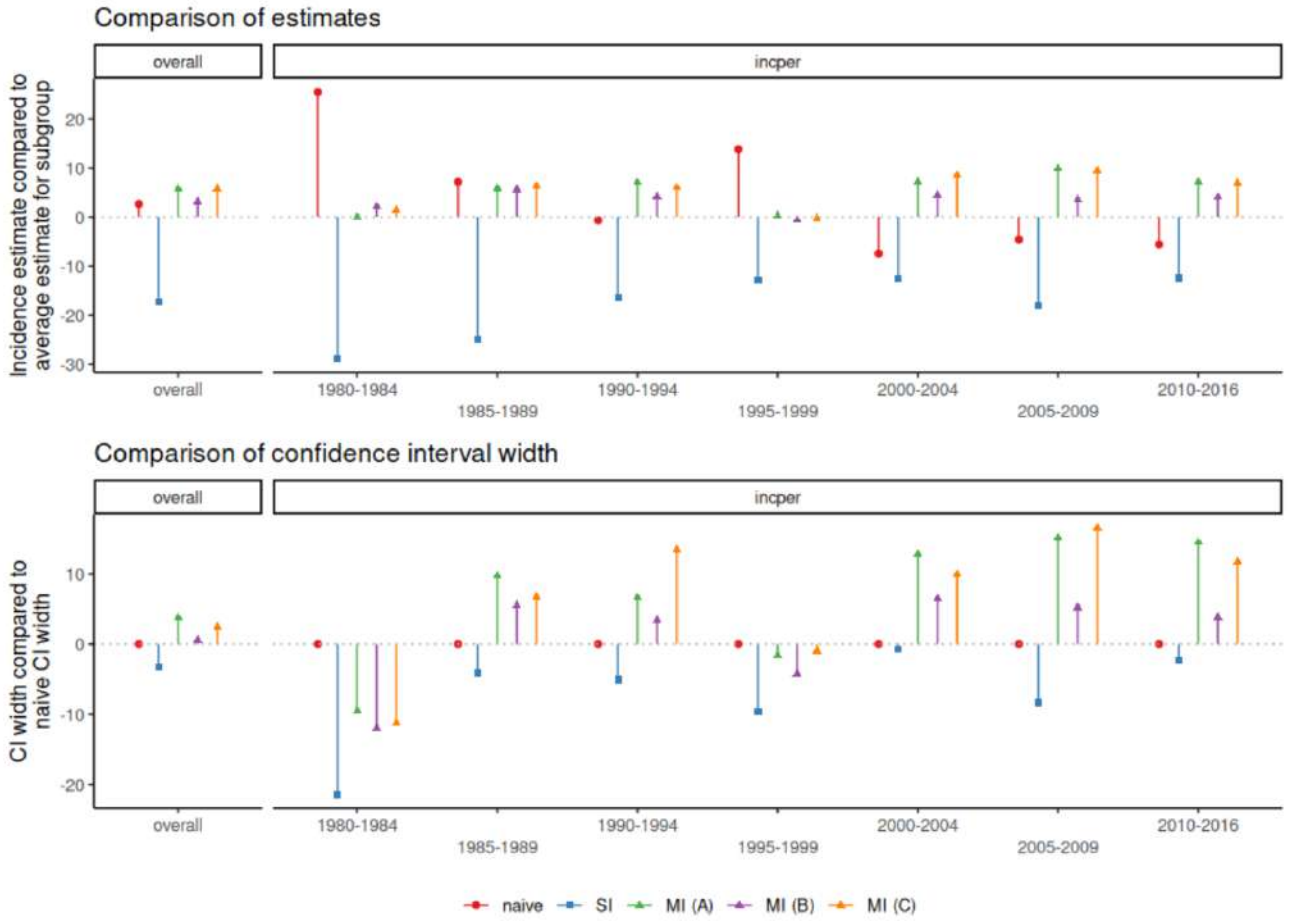
N = 24,612 (n = 1264 [5%] were lost to follow-up)

Method	**details**
Naive (NI)	*(no change to person-years or vital status)*
Single imputation (SI)	Follow-up time “simulated” 1x with average hazard rate of 0.0591 per year (based on Swiss-wide data for breast cancer)
Multiple Imputation (MI)	Imputed 25 times using [...], analyzed and estimated incidence rates pooled
A	~ age group + incidence period + morphology + baseline hazard
B	~ age + age^2 + incidence year + incidence year^2 + morphology
C	~ age + incidence year + morphology

Results


overall
incidence period
estimated CBC incidence (per 100,000 person−years)
800
600
400
200
0
393
589
238
naive
SI
MI (A)
MI (B)
MI (C)
overall
1980−1984
1985−1989
1990−1994
1995−1999
2000−2004
2005−2009
2010−2016

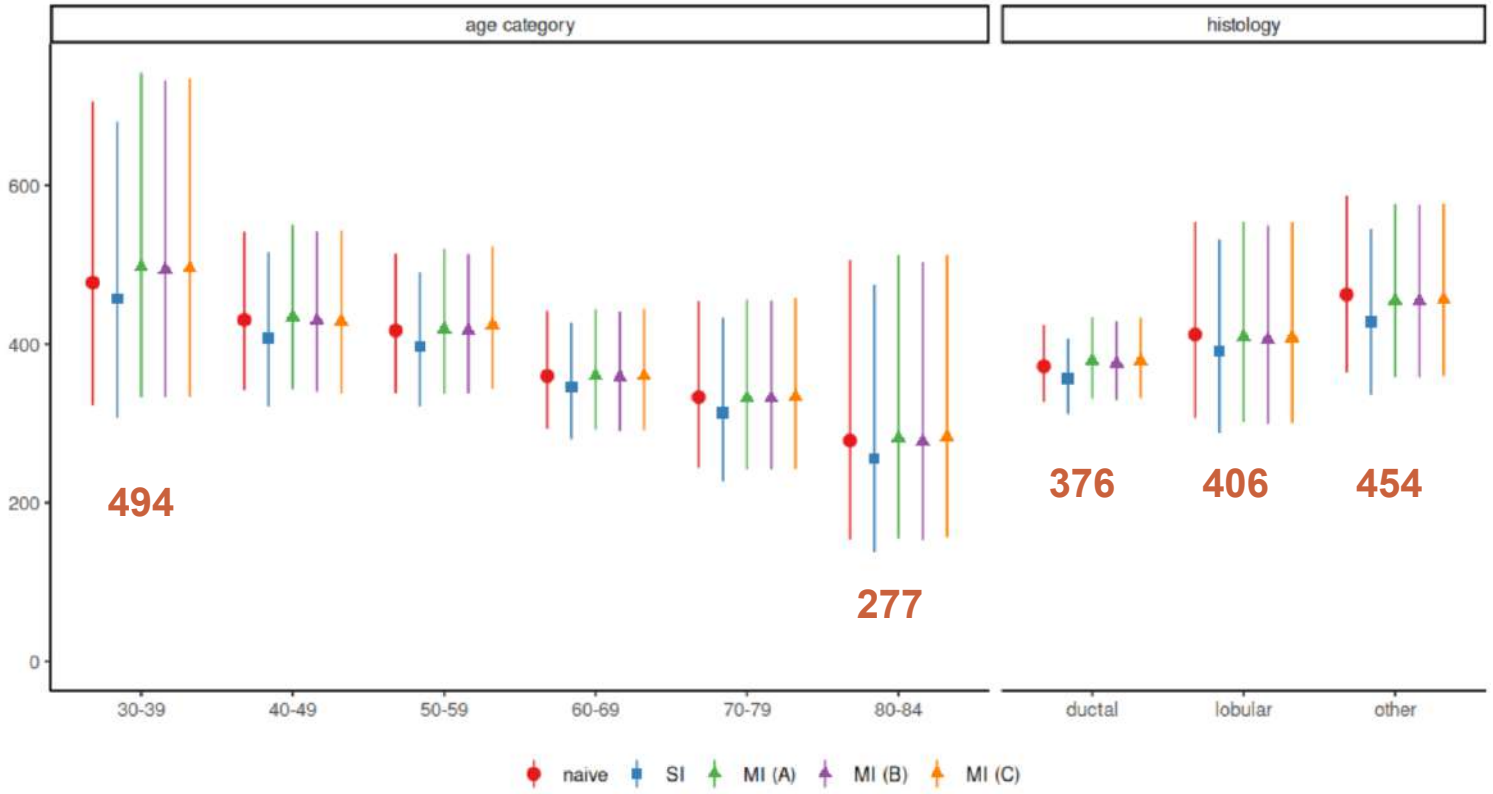
Results


Comparison of estimates
overall
incper
Incidence estimate compared to average estimate for subgroup
20
10
0
-10
-20
-30
overall
1980-1984
1985-1989
1990-1994
1995-1999
2000-2004
2005-2009
2010-2016
Comparison of confidence interval width
overall
incper
CI width compared to naive CI width
10
0
-10
-20
overall
1980-1984
1985-1989
1990-1994
1995-1999
2000-2004
2005-2009
2010-2016
naive
SI
MI (A)
MI (B)
MI (C)

Conclusions

- CBC incidence continues to decline
- Different approaches to handle missing follow-up time or vital status gave similar results, but:
  - SI results were always lower with narrower confidence intervals
  - MI results were similar to naive approach, but had slightly wider confidence intervals
- Multiple imputation enables inclusion of all cancer registry subjects
- MI can also handle missing covariates

Results


age category
histology
600
400
200
0
494
277
376
406
454
30-39
40-49
50-59
60-69
70-79
80-84
ductal
lobular
other
naive
SI
MI (A)
MI (B)
MI (C)