Extending standard reporting to improve communication of survival statistics

Tor Åge Myklebust

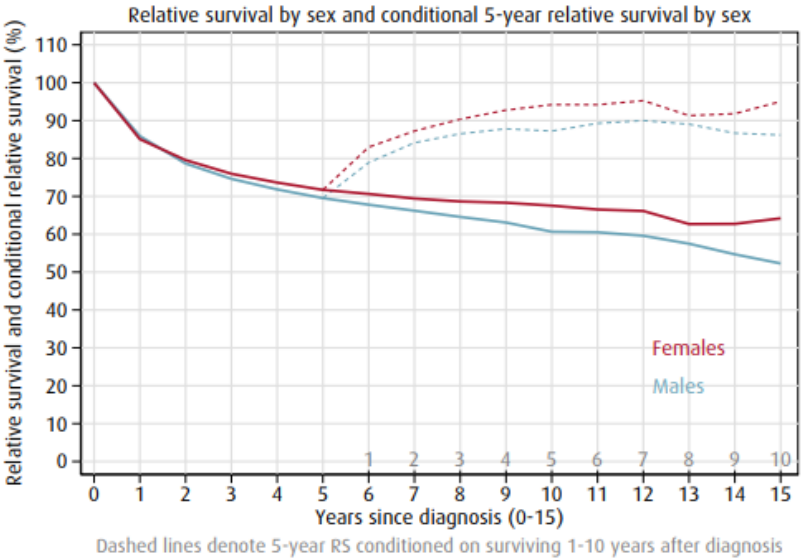
Cancer Registry of Norway, Norwegian Institute of Public Health

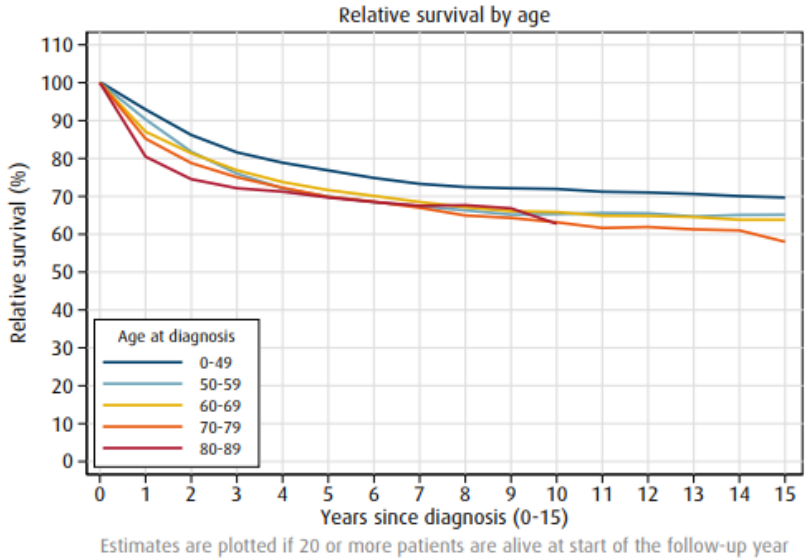
Background

- Relative/net survival estimates is the most widely used cancer survival measure in routine reports/publications.
- Suitable for comparisons as it removes effects of differential 'other cause' mortality.
- Interpretation is not straightforward: *The probability of surviving the cancer of interest in a world where you cannot die of anything else*.
- And so might not be so relevant for communicating prognosis to patients, clinicians etc.

Background

Figure 8.2–E: Colon (ICD-10 C18)





Background

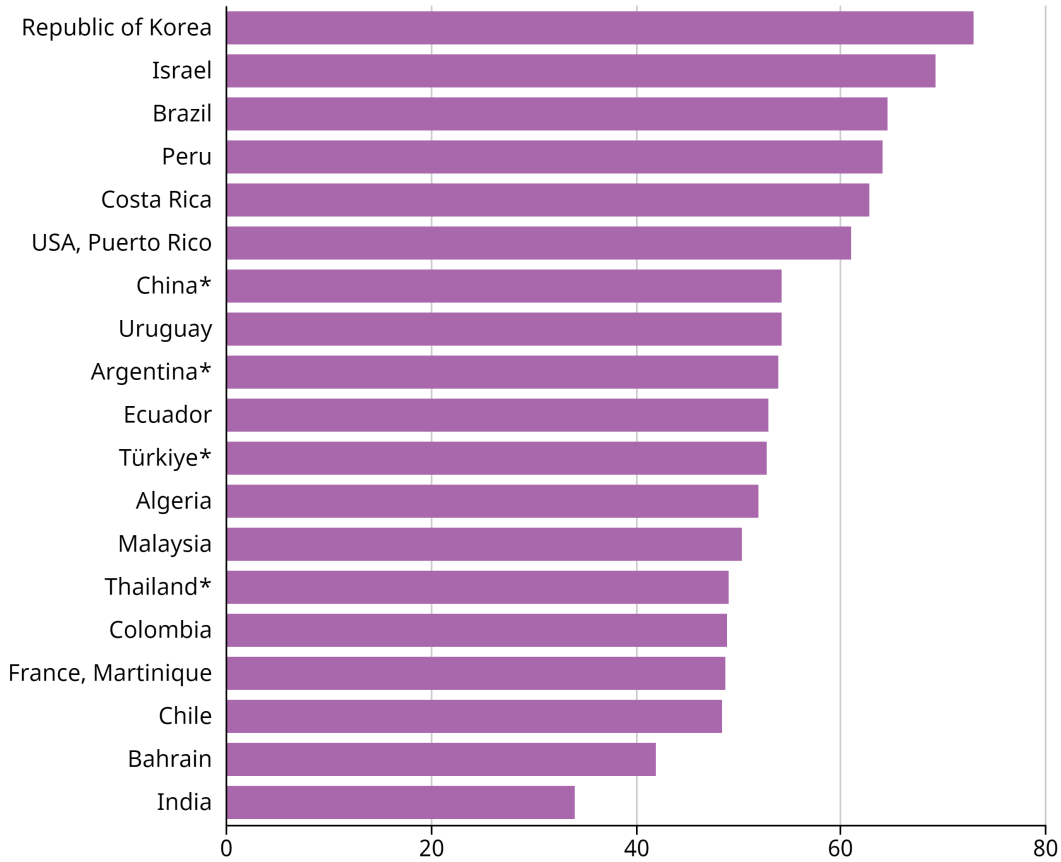
Cancer in Norway 2021

Table 8.1: Five-year relative survival by primary site, stage and period of diagnosis, 1982–2021, males

			Relative survival (%)							
ICD-10	Site	Stage	1982–86	1987–91	1992–96	1997–01	2002–06	2007–11	2012–16	2017–21*
C00–96	All sites	Total	43.7	46.8	51.9	57.6	63.1	69.4	74.4	77.1
C00–14	Mouth, pharynx	Total	56.6	58.1	55.5	55.1	55.9	65.0	67.9	72.5
		Localised	76.3	81.0	81.6	81.1	80.4	83.1	85.1	87.7
		Regional	25.9	27.4	27.3	33.4	39.2	50.4	61.8	65.2
		Distant	-	7.3	11.8	6.2	12.6	7.8	6.1	10.4
		Unknown	54.8	38.2	51.9	56.2	60.7	76.0	57.4	72.7

Background

Net survival based on ICSS standards, 5-year, both sexes, cases diagnosed 2008–2012
Colon, Worldwide, Net survival based on ICSS standards
** Median survival estimate for the country*



Background

- The goal of this project was to:
  1. Estimate and present alternative survival measures across a large range of cancer types.
  2. To propose a way of automizing such statistics.

Survival measures

1. Cumulative incidence/Crude probabilities
2. Expected remaining lifetime
3. Lifeyears lost
4. Reference-adjusted survival measures
5. Conditional survival

Methods

- Estimated flexible parametric survival models in a relative survival framework (stpm2/stpm3, Paul Lambert)
- Using a 5-year period window for 'up-to-date' estimates
- Maximum follow-up set to 15 years
- Estimated models separately by cancer site
- Estimated different models for assessing trends and for producing 'up-to-date' estimates/predictions relevant for recently diagnosed patients

Methods

- Covariates included age at diagnosis (splines), SEER summary stage (not for trends) and sex
- Stage missing to a varying degree → imputed (mi impute...)
- Models were estimated separately on each complete dataset and predictions combined across datasets using Rubins' rules

Predictions

Also choices to be made for how to generate predictions from the estimated models

- For recently diagnosed the main aim was to calculate measures most relevant to certain patient groups
- We make individual predictions for each patient in that group (with their covariate values)
- Take the average of individual predictions for each group (Standsurv, Paul Lambert)
- Estimates are not comparable across groups
- When calculating predictions for conditional measures we predicted for median aged patients

- For trends we predicted using the age distribution in the last 5-year period

Model selection

How to automate model selection?

- We have 23 different cancer sites
- They vary in size, prognosis, age distribution, stage distribution
- Convergence problems are unavoidable (at least on Norwegian data)
- We chose to pre-specify a 'menu' of different models
- Starting with the most desirable model (we think) at the top
- ...and gradually simplifying model specifications by
  1. Reducing DF used for modelling TVCs, baseline EH and/or age
  2. Removing interactions between covariates
  3. Removing interaction with follow-up (i.e. assuming proportionality)
- A total of 20 different models

Model selection

For 'up-to-date predictions' the first model on the menu is something like

```
stpm2 sex rcs_age1-rcs_age4 stage2-stage3 sex#stage, ///
df(5) tvc(rcs_age1-rcs_age4 stage2-stage3) ///
dftvc(3) scale(hazard) bhazard(rate)
```

whereas the last model on the menu is something like

```
stpm2 sex rcs_age1-rcs_age2 stage2-stage3, ///
df(3) scale(hazard) bhazard(rate)
```

Speeding up calculations (Thanks to Bjarte Aagnes)

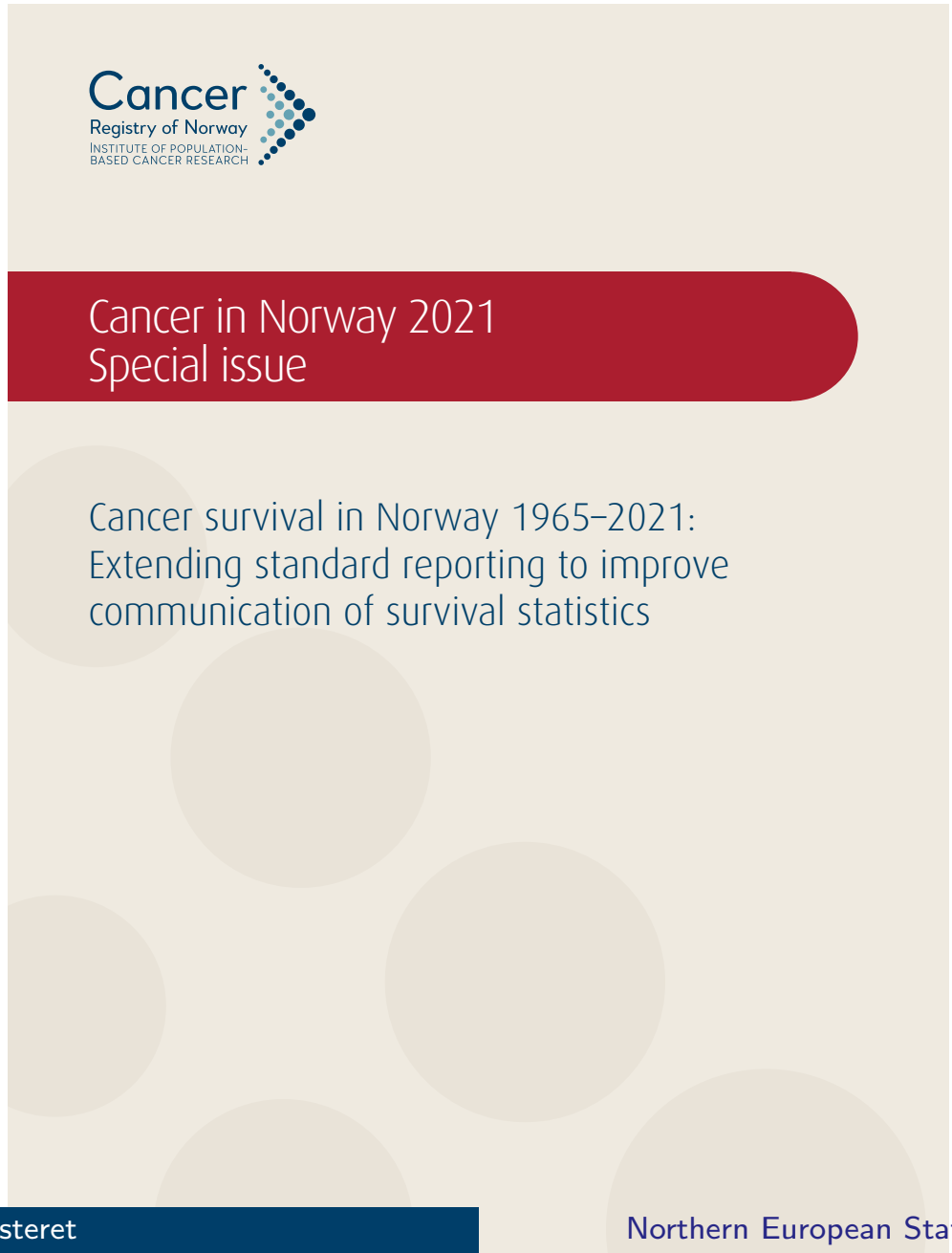
- Used a split-apply-combine strategy running parallell Stata sessions on local machine
- Avoid exhausting resources
  - 24 processors and 32 GB RAM
  - Used sysresources[1] to check CPU-load and available free memory
  - Starting new session if:
    1. CPU-load < 75
    2. Free memory > 25

[1] https://github.com/wbuchanan/StataOS

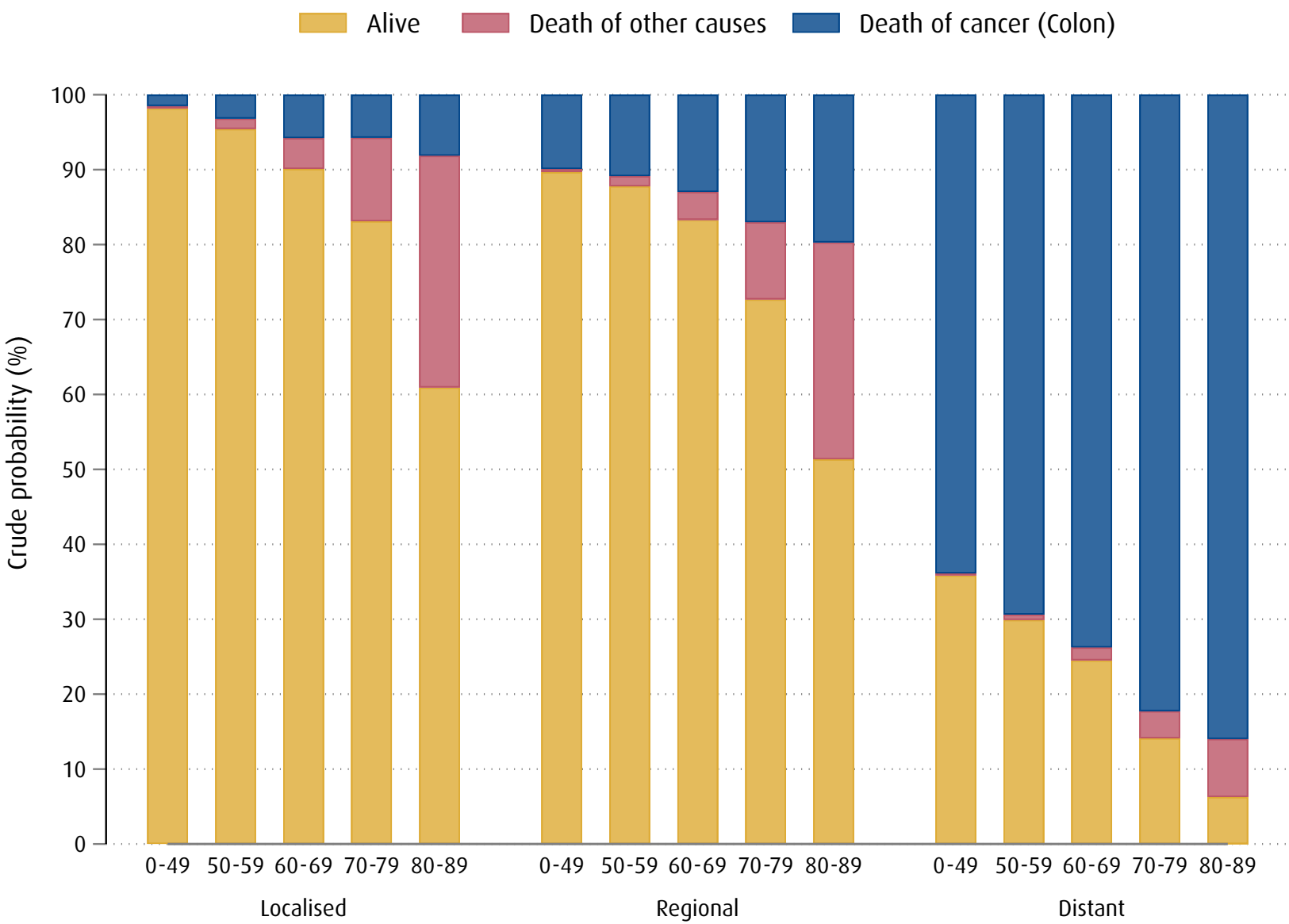
Speeding up calculations (Thanks to Bjarte Aagnes)

```
forvalues i=1(1)$N_imputations {
        sleep `=10*1000'
        while 1 {
            sysresources
            if ( r(pctfreemem) < 25 | r(cpuload) > 0.75 ) {
                sleep `=30*1000'
                continue
            }
            else {
                winexec $StataExe /e do `dofile' `args'
                continue, break
            }
        }
    }
```

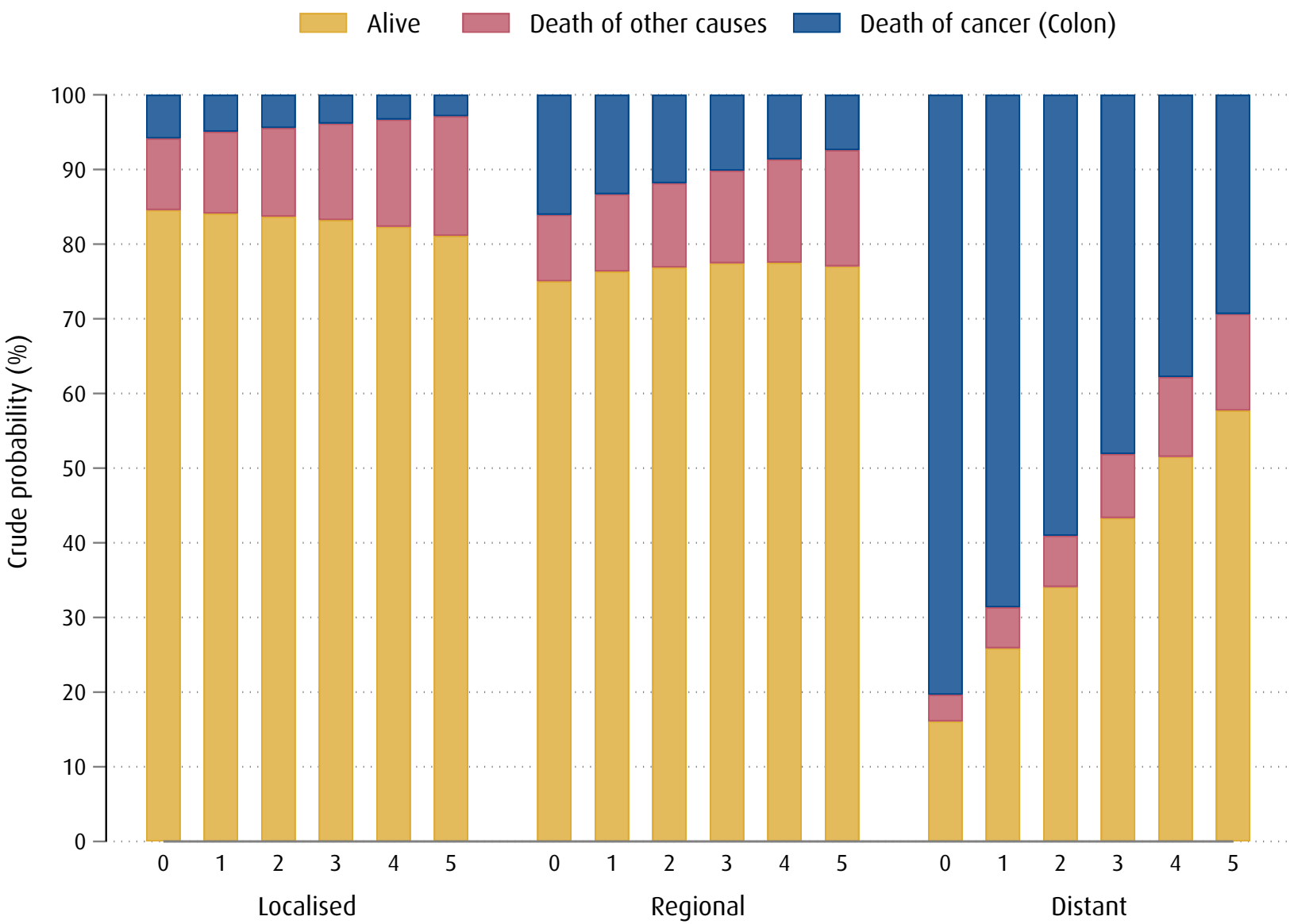
Results for 23 cancer sites



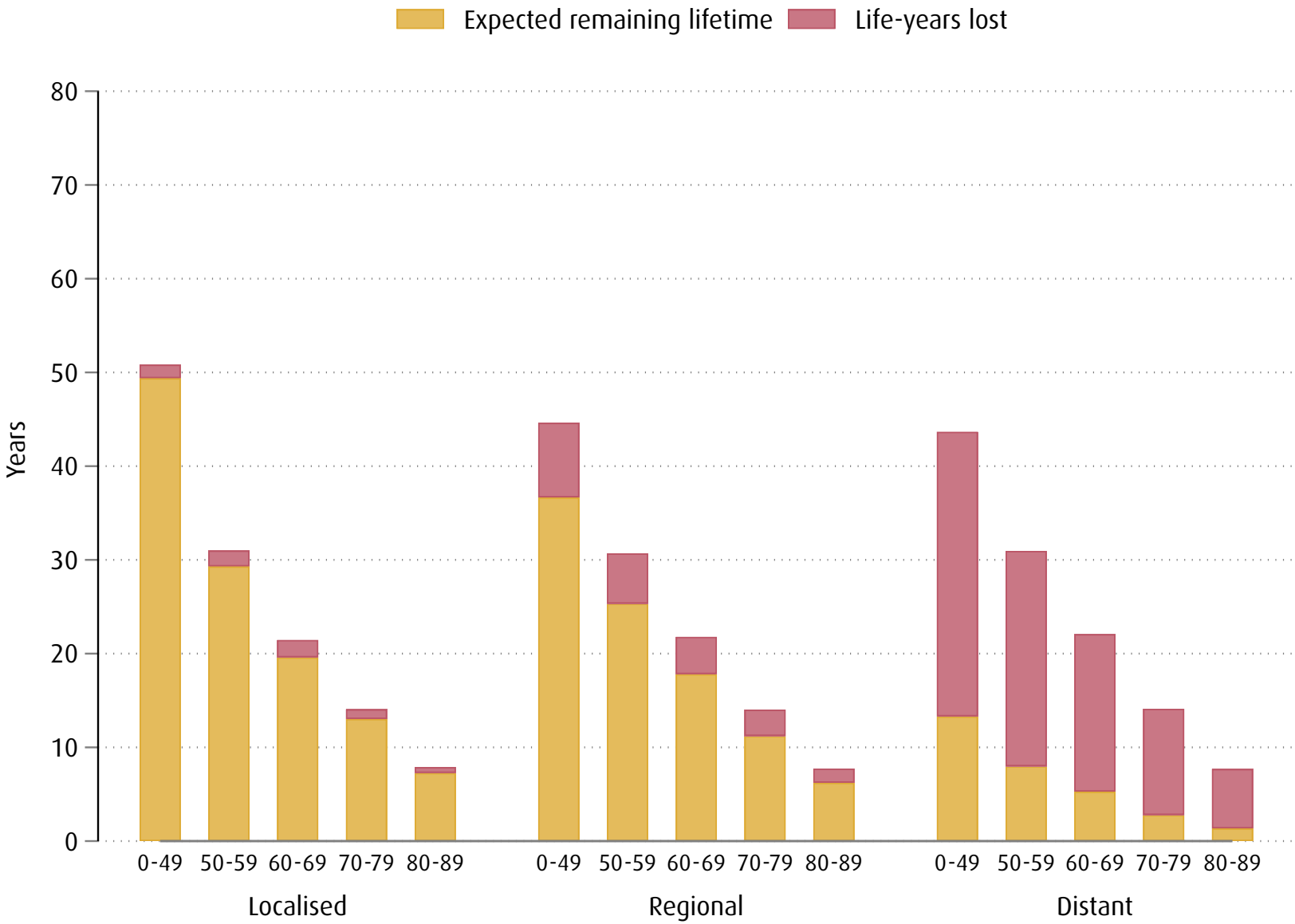
Results colon cancer women



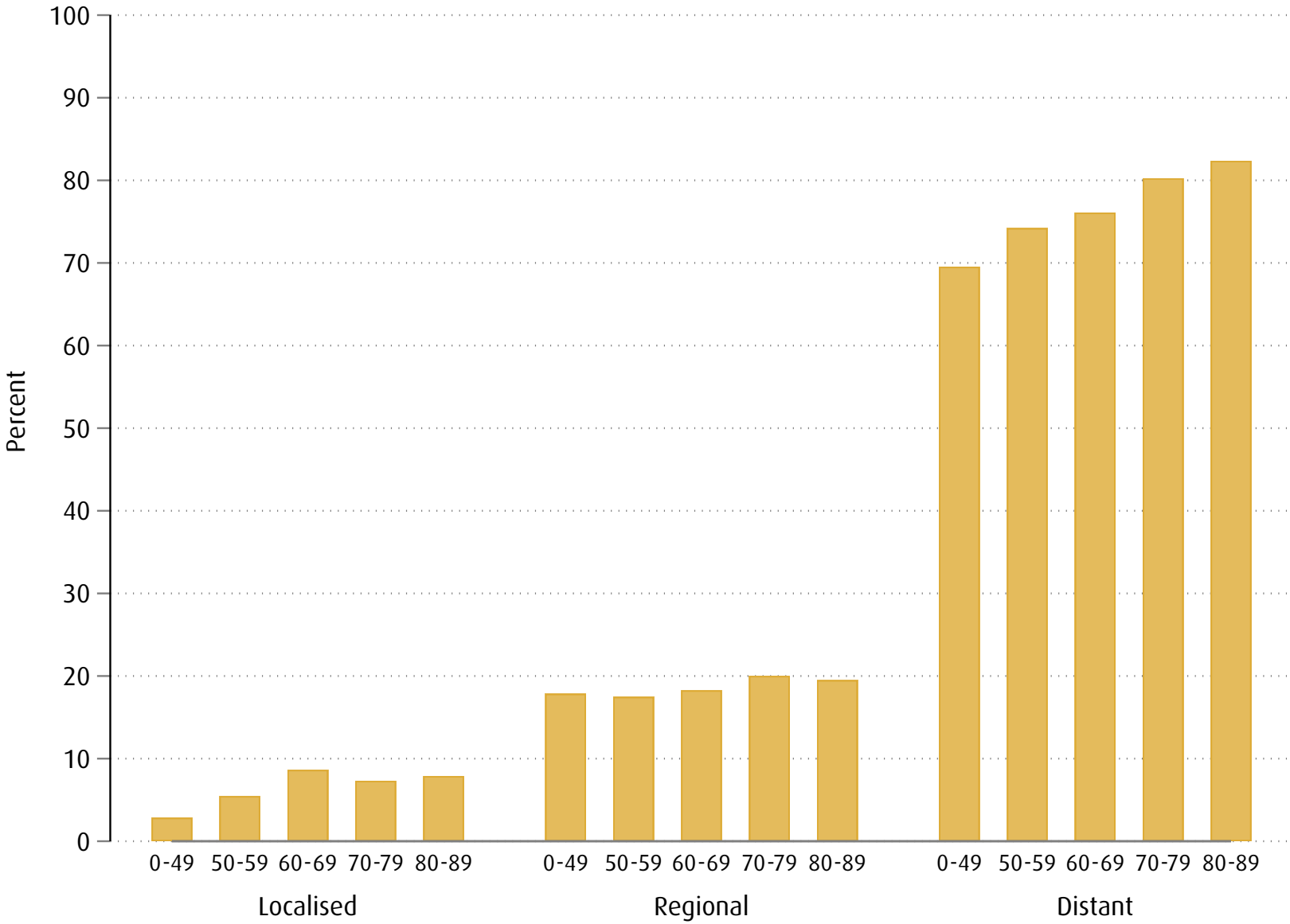
Results colon cancer women



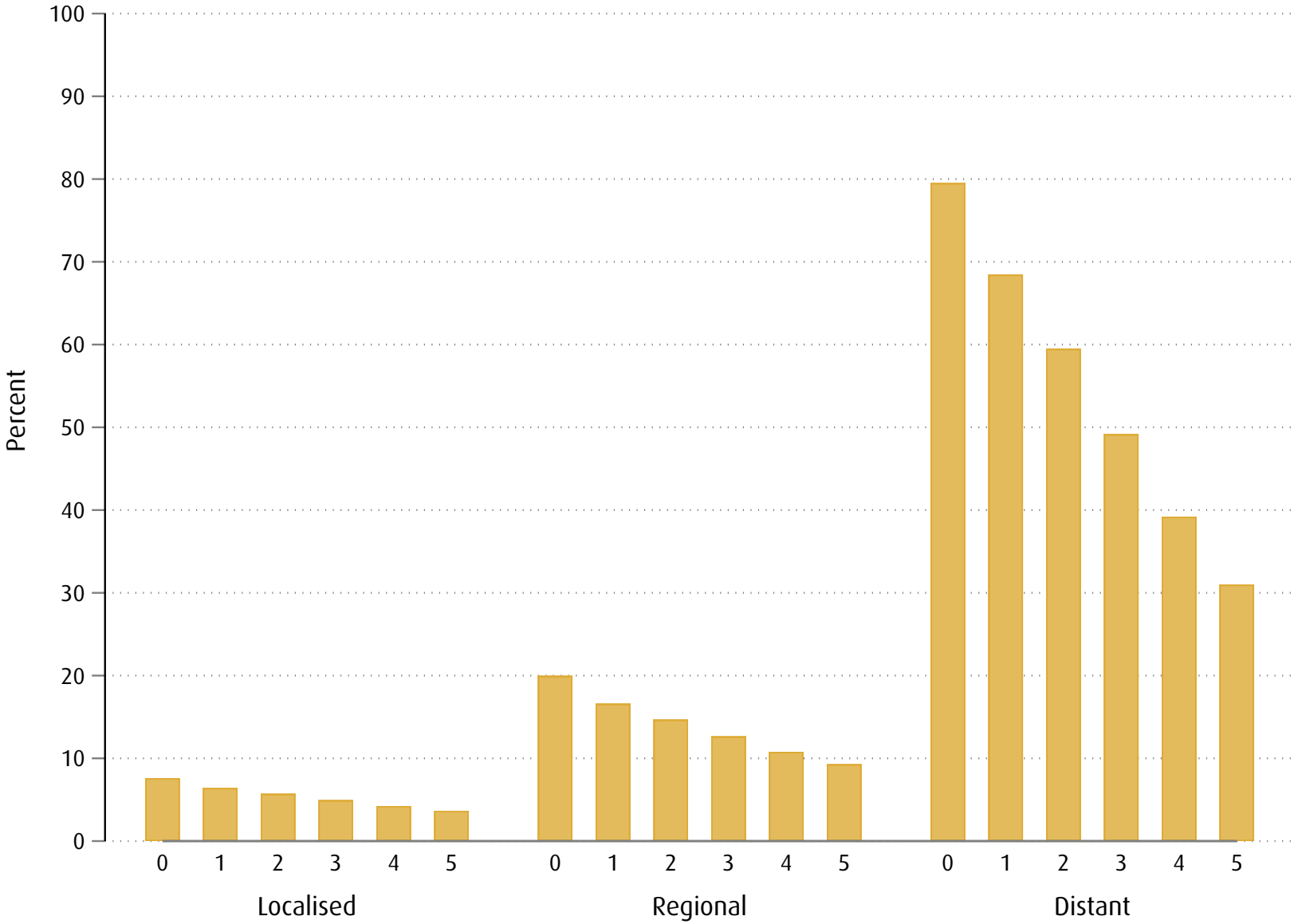
Results colon cancer women



Results colon cancer women



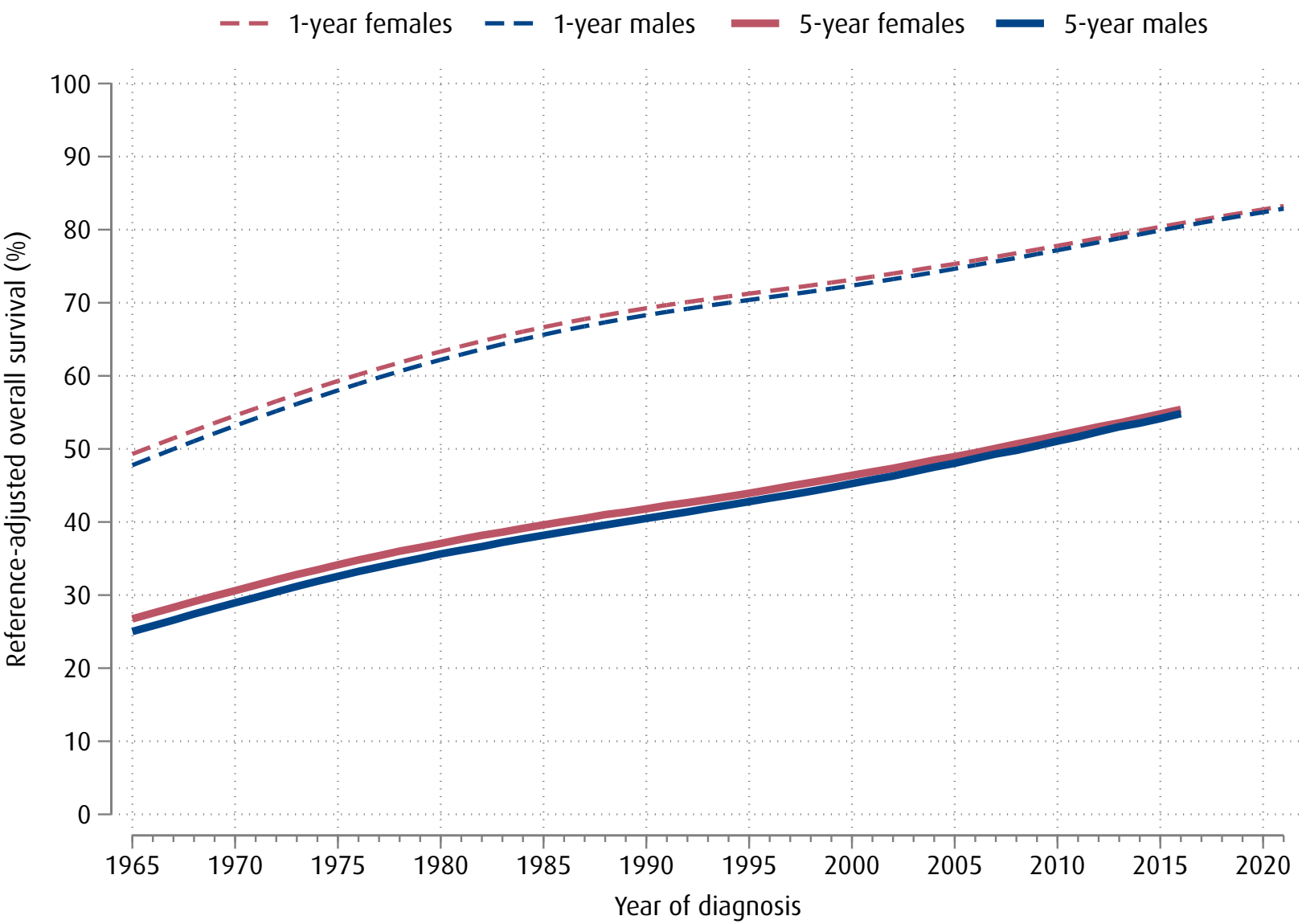
Results colon cancer women



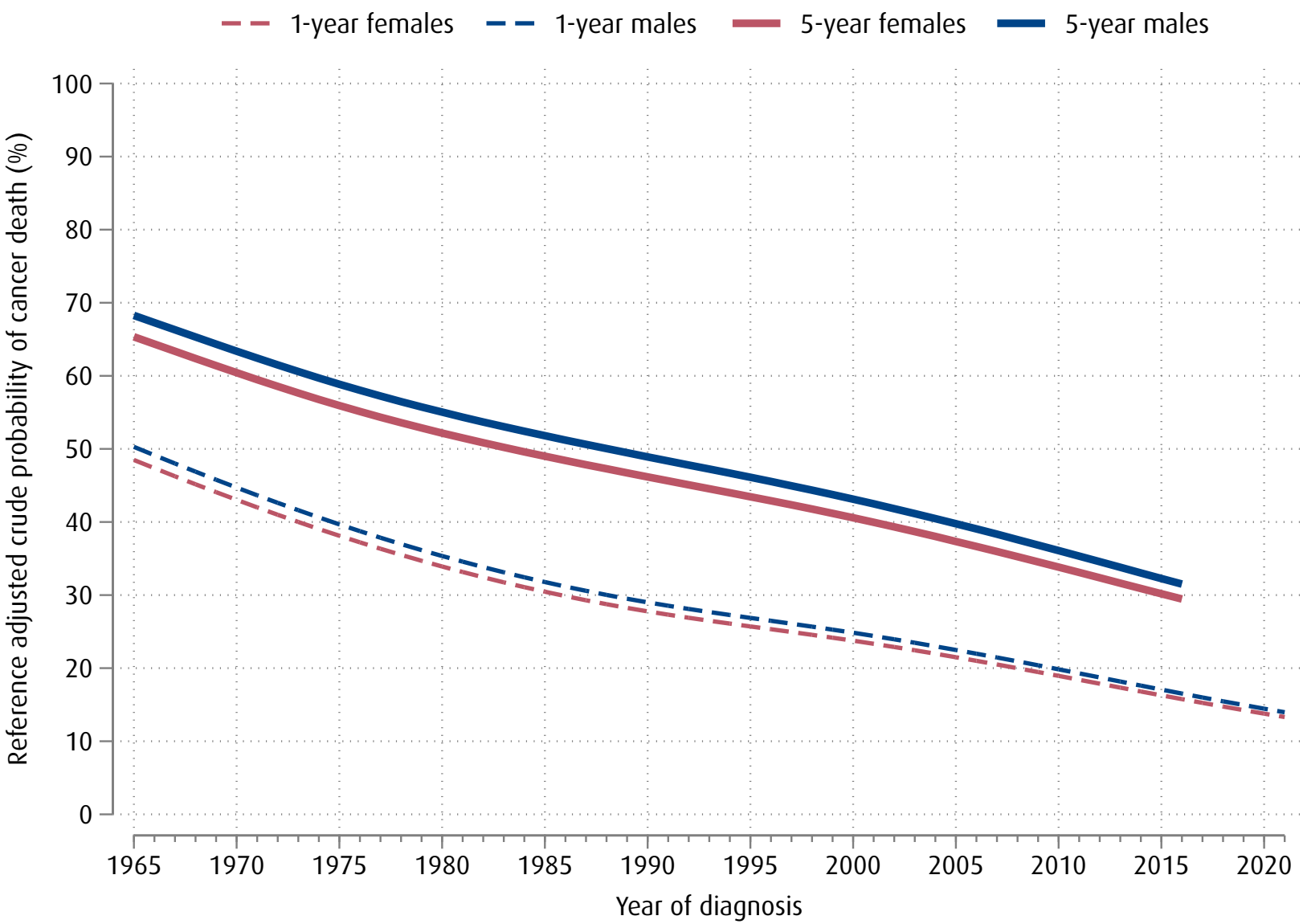
Reference-adjusted measures

- When estimating relative survival we effectively set other cause mortality to zero
- Not necessarily the most logical choice
- We could instead fix the level of other cause mortality to a reference
- Apply mortality rates for 2021 backwards in time
- Use Norwegian population mortality rates to Swedish cancer patient data

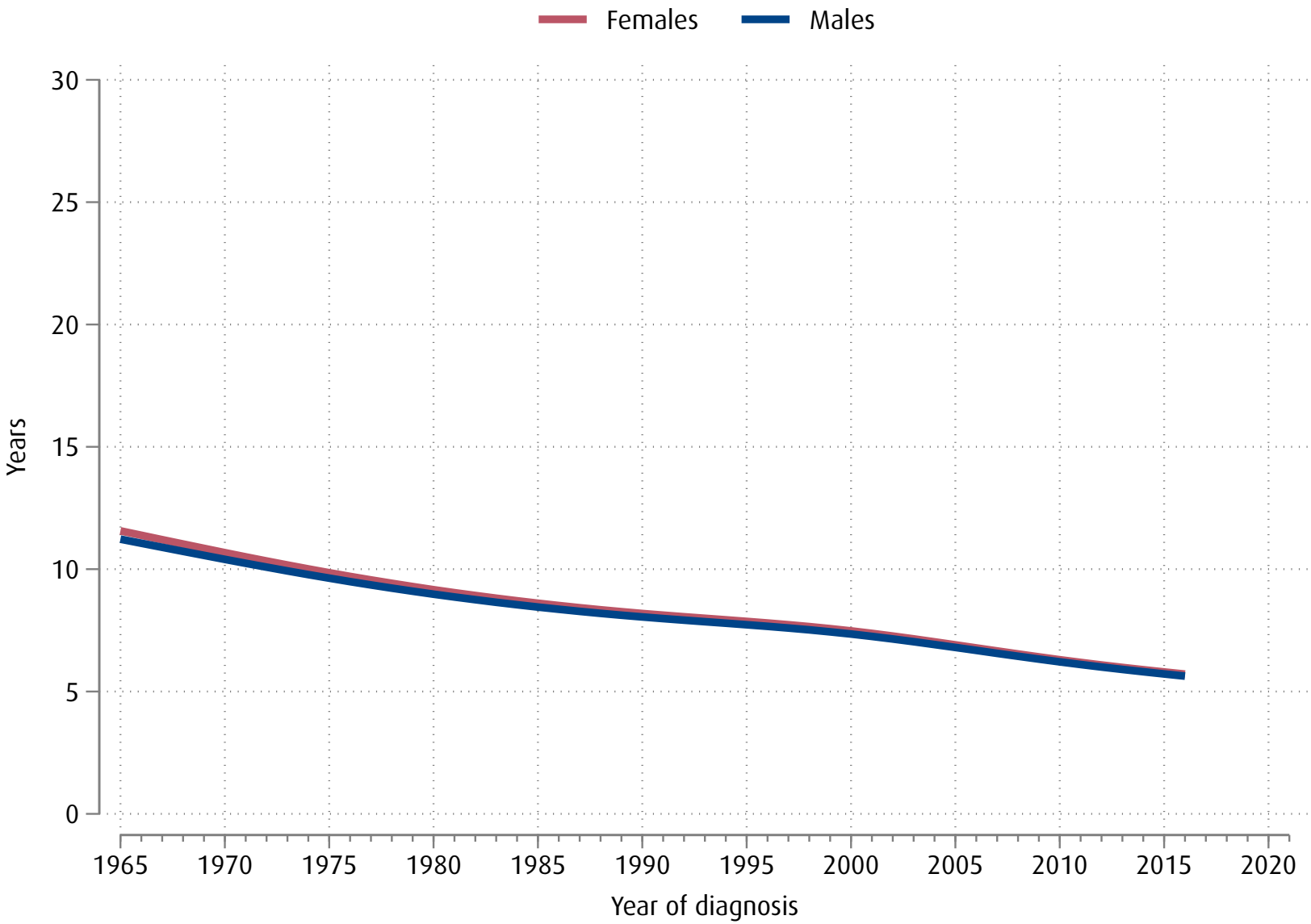
Overall survival colon cancer



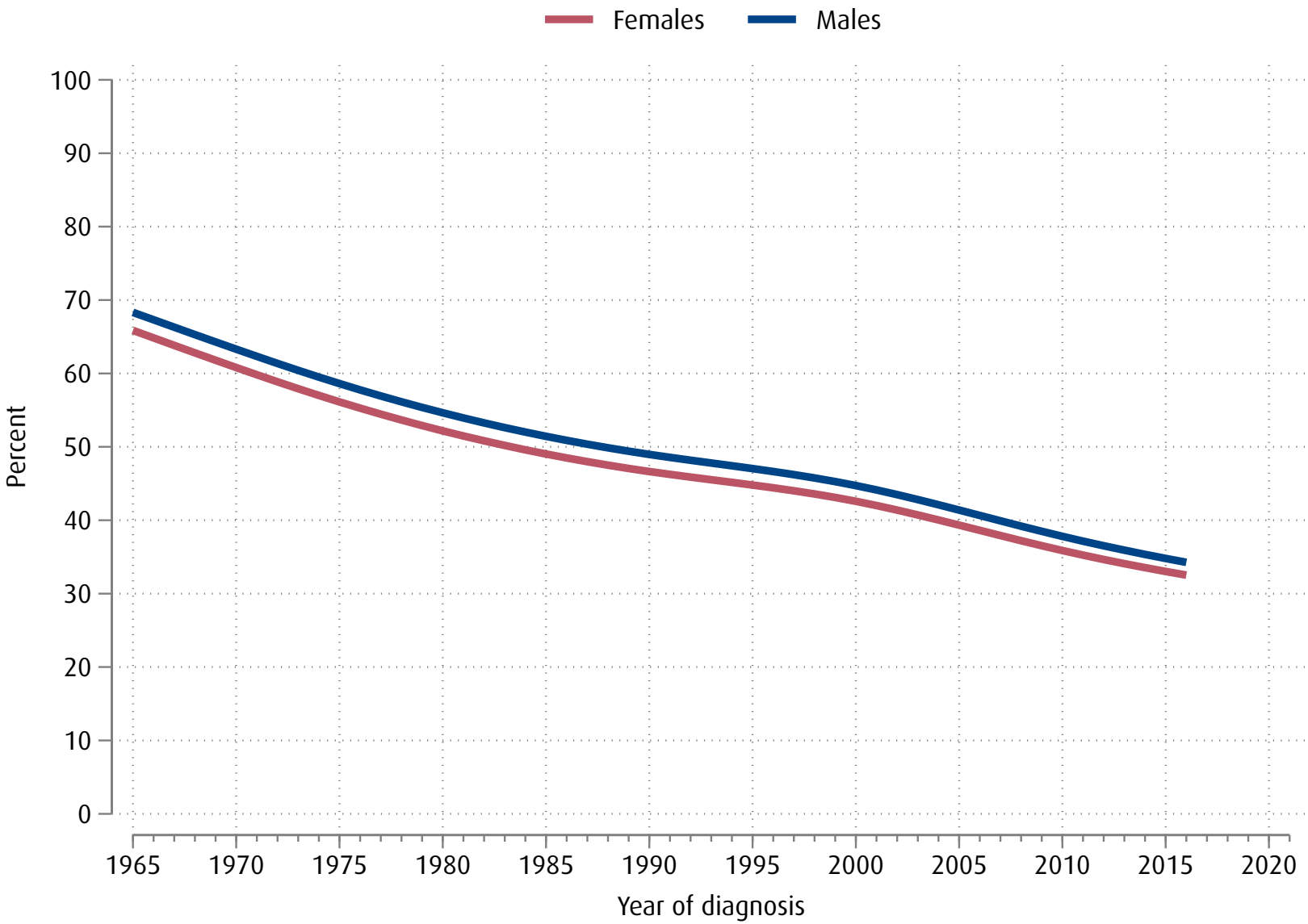
Crude probabilities colon cancer



Life years lost colon cancer



Life years lost colon cancer



Summary

- Standard reporting of cancer survival statistics should include survival measures that are aimed towards quantifying prognosis
- More and better quality registries will enable us to make statistics that are even more clinically relevant than today
- CRN should be the primary source of information regarding cancer prognosis

Project group

- Bjarte Aagnes
- Yngvar Nilssen
- Bjørn Møller
- Paul W. Dickman
- Paul C. Lambert
- Therese M.L. Andersson
- Anna L.V. Johansson
- Mark Rutherford

Documentation