Session 19
Crude probability of death: estimation and application

Paul W Dickman¹ and Paul C Lambert¹,²

¹Department of Medical Epidemiology and Biostatistics,
Karolinska Institutet, Stockholm, Sweden
²Department of Health Sciences,
University of Leicester, UK

Cancer survival: principles, methods and analysis
LSHTM
June 2014
Overview

- Introduction to competing risks.
- Net survival versus crude survival; concepts and definitions.
- Which measure (crude or net) is most relevant for my research question?
- Assumptions and estimation.
- My presentation will focus on concepts; I have included slides with details and examples that I won’t cover.

Survival of patients with colon cancer in Finland

- Coding of vital status
  
<table>
<thead>
<tr>
<th>Freq.</th>
<th>Numeric</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>4642</td>
<td>0</td>
<td>Alive</td>
</tr>
<tr>
<td>8369</td>
<td>1</td>
<td>Dead: colon cancer</td>
</tr>
<tr>
<td>2549</td>
<td>2</td>
<td>Dead: other</td>
</tr>
</tbody>
</table>

- The event of interest is death due to (colon) cancer.
- Other events are known as ‘competing events’ or ‘competing risks’.
- Based on the research question, we choose between one of two quantities to estimate:
  - Eliminate the competing events (estimate net survival)
  - Accommodate the competing events (estimate crude survival)

Many synonyms for the same concept

- Net probability of death due to cancer = Probability of death in a hypothetical world where the cancer under study is the only possible cause of death
- Crude probability of death due to cancer = Probability of death in the real world where you may die of other causes before the cancer kills you

- Net probability also known as the marginal probability.
- Crude probability also known as the cause-specific cumulative incidence function (Geskus) or the cumulative incidence function.
Net survival

- In cancer patient survival we typically choose to eliminate the competing events.
- That is, we aim to estimate net survival (using either cause-specific survival or relative survival).
- It is important to recognise that net survival is interpreted in a hypothetical world where competing risks are assumed to be eliminated. Requires conditional independence.
- I will later show how ‘real world’ probabilities (i.e., crude probabilities) can be estimated by accommodating, rather than eliminating, competing risks.

The choice between relative and cause-specific survival for estimating net survival

- Both methods involve assumptions specific to the approach:
  - **Cause-specific** Accurate classification of cause-of-death
  - **Relative** Appropriate estimation of expected survival
- We choose the approach for which we have the strongest belief in the underlying assumptions.
- For population-based studies this is typically relative survival but every study must be evaluated on its specific merits.
Cancer death (status=1) as the outcome

use colon if age > 70
stset exit, origin(dx) failure(status==1) scale(365.25)
sts graph, fail

Non-cancer death (status=2) as the outcome

use colon if age > 70
stset exit, origin(dx) failure(status==2) scale(365.25)
sts graph, fail
Net probabilities do not sum to the total probability of death!

Two estimates of the same quantity (net survival)

But they are very similar for ‘young’ patients
Why the hypothetical world?

- Net survival estimates survival in the hypothetical world where you cannot die of causes other than the cancer of interest.
- Net survival is a theoretical construct. We can attempt to estimate it using either cause-specific survival or relative survival.
- It is useful as a measure of cancer-patient survival that is independent of background mortality so we can make comparisons across time, across different age-groups in our population and across different countries.

Interpreting estimates of net survival

- The cumulative relative survival ratio can be interpreted as the proportion of patients alive after $i$ years of follow-up in the hypothetical situation where the cancer in question is the only possible cause of death.
- $1 - \text{RSR}$ can be interpreted as the proportion of patients who will die of cancer within $i$ years of follow-up in the hypothetical situation where the cancer in question is the only possible cause of death.
- We do not live in this hypothetical world (where we estimate what is called the net probability of death). Estimates of the proportion of patients who will die of cancer in the presence of competing risks can also be made (crude probabilities of death).
- Cronin and Feuer (2000) [1] extended the theory of competing risks to relative survival; their method is implemented in our Stata command `strs`.

Net (left) and crude (right) probabilities of death in men with localized prostate cancer aged 70+ at diagnosis (Cronin and Feuer [1])
What this means

- Among these men, the probability of dying of prostate cancer within 15 of diagnosis is 40% in the hypothetical world where it is not possible to die of other causes. This is how the relative survival or cause-specific survival is interpreted.
- However, in the real world where it is possible to die of other causes the probability of dying of prostate cancer within 15 years of diagnosis is less than 20%.

What this means

- In the hypothetical world, where it is not possible to die of causes other than breast cancer, the probability of dying of regional breast cancer is similar for all age groups.
- However, in the real world the probability of dying of breast cancer is lower for elderly women because they are more likely to die of other causes.
Relative survival was estimated to be 50%.

### Cancer survival statistics

- 50% of adult cancer patients diagnosed in 2010-2011 in England and Wales are predicted to survive 10 or more years.
- 46% of men and 54% of women cancer patients diagnosed in 2010-2011 in England and Wales are predicted to survive 10 or more years.
- Cancer survival rates in the UK have doubled in the last 40 years.

---

**Should we estimate crude or net survival?**

**Objective:** To assess the association between β-blockers and PCa-specific mortality in a cohort of 3561 prostate cancer patients with high-risk or metastatic disease, and to address potential confounding from the use of statins or acetylsalicylic acid (ASA).

**Design, setting, and participants:** Clinical information from all men reported to the Cancer Registry of Norway with a PCa diagnosis between 2004 and 2009 (n = 24,571) was coupled with information on filled prescriptions between 2004 and 2011 from the Norwegian Prescription Database. Exclusion criteria were low- or intermediate-risk disease; planned radiotherapy or radical prostatectomy; initiation of β-blocker, ASA, or statin use after diagnosis where applicable; missing information on baseline Gleason score, prostate-specific antigen level, T stage or performance status; and missing follow-up.

---

**Should we estimate crude or net survival?**

- We wish to compare PrCa-specific survival among two groups:
  - Men with prostate cancer using beta-blockers
  - Men with prostate cancer not using beta-blockers

- The authors used a ‘competing risks analysis’ and concluded that the men who used beta-blockers were less likely to die of prostate cancer.

- Bernard co-authored a commentary that very nicely explained why such an analysis is incorrect [2] and that net probabilities are more appropriate than crude probabilities.
**ABSTRACT**

The need to develop treatments and/or programs specific to a disease requires the analysis of outcomes to be specific to that disease. Such endpoints as heart failure, death due to a specific disease, or control of local disease in cancer may become impossible to observe due to a prior occurrence of a different type of event (such as death from another cause). The event which hinders or changes the possibility of observing the event of interest is called a competing risk.

The usual techniques for time-to-event analysis applied in the presence of competing risks give biased or uninterpretable results. The estimation of the probability of the event therefore needs to be calculated using specific techniques such as the cumulative incidence function introduced by Kalbfleisch and Prentice. The model introduced by Fine and Gray can be applied to test a covariate when competing risks are present. Using specific techniques for the analysis of competing risks will ensure that the results are unbiased and can be correctly interpreted.

---

**This article is misleading and causes confusion**

- ‘When the CR are ignored and the CR observations are censored the analysis reduces to a ‘usual’ time-to-event scenario. Due to the familiarity of this type of analysis and the availability of software, many researchers resort to this approach, as seen in the earlier examples. However, it is unanimously agreed not only among statisticians, that the estimation of the probability of event in this case overestimates the true probability.’

---

**Net (left) and crude (right) probabilities of death in men with localized prostate cancer aged 70+ at diagnosis (Cronin and Feuer [1])**
Natural frequencies presented using infographics

- number who will likely die from their cancer
- number who will likely die from other health related causes
- number who will likely survive

1 Year After Diagnosis
5 Years After Diagnosis

Cancer Survival Query System (Rocky Feuer)

It is estimated that by:
1 year after diagnosis:
Approximately 2 out of 100 will die from their cancer,
Approximately 10 out of 100 will die from other causes,
Approximately 88 out of 100 will survive.

5 years after diagnosis:
Approximately 12 out of 100 will die from their cancer,
Approximately 47 out of 100 will die from other causes,
Approximately 41 out of 100 will survive.

The independence assumption - crucial for the interpretation of survival curves when competing risks are present

Independence assumption
The time to death from the cancer in question is conditionally independent of the time to death from other causes. i.e., there should be no factors that influence both cancer and non-cancer mortality other than those factors that have been controlled for in the estimation.
If the independence assumption is not satisfied

- Cause-specific survival curves provide biased estimates of net survival.
- If it is not possible to adjust for the mechanism that introduces the dependence then survival curves should be interpreted with care.
- However, the cause-specific hazard rates still have a useful interpretation as the rates that are observed when competing risks are present.

If the independence assumption is satisfied

- Cause-specific survival curves provide estimates of net survival (provided that the classification of cause-of-death is accurate).
- The survival curves are interpreted as the survival that we would observe if it was possible to eliminate all competing causes of death.
- This is a strictly hypothetical (but useful!) construct.
- The cause-specific hazard rates provide estimates of the rates that we would observe in the absence of competing causes of death.
- In the competing risks literature, net survival and hazard are typically referred to as marginal survival and hazard, respectively.

How can we test if the independence assumption is satisfied?

Simple answer

You can't!

- It is not possible to formally test if the independence assumption is satisfied based on the observable data.
- You have to make the decision of whether your estimates provide an estimate of something useful based on your subject matter knowledge.
How do these concepts translate to relative survival?

- Cause-specific survival and relative survival aim to estimate the same underlying quantity.
- Even though we do not make use of explicit cause of death information, the independence assumption applies also in a relative survival framework.
- If satisfied, relative survival and excess mortality rates provide estimates of net survival (given that the patients are exchangeable to the population used to estimate expected survival)
- If not satisfied, relative survival curves are biased whereas the excess mortality rates are interpretable as real-world rates (in the presence of competing risks).

Estimating crude probabilities (life table, RS)

- The contribution to the crude probability for each interval.

Interval crude probability of death due to cancer

\[ \hat{g}_{jc} = \left( \prod_{i=1}^{j-1} \hat{p}_{j} \right) \left( 1 - \frac{\hat{p}_{j}}{\hat{p}_{j}^*} \right) \left( 1 - \frac{1}{2} \left( 1 - \hat{p}_{j}^* \right) \right) \]

Interval crude probability of death due to other causes

\[ \hat{g}_{jo} = \left( \prod_{i=1}^{j-1} \hat{p}_{j} \right) \left( 1 - \hat{p}_{j}^* \right) \left( 1 - \frac{1}{2} \left( 1 - \frac{\hat{p}_{j}}{\hat{p}_{j}^*} \right) \right) \]

- First term: survival to start of interval.
- Second term: probability of dying of cause in interval.
- Third term: needed as dealing with grouped time.

Estimation II

- Cumulative crude probabilities obtained by summing interval estimates.

\[ C_{rc} = \sum_{i=1}^{j} \hat{g}_{jc} \]

\[ C_{ro} = \sum_{i=1}^{j} \hat{g}_{jo} \]

- \( C_{rc} + C_{ro} \) gives the all-cause probability of death at the end of interval \( j \)
- Also possible to obtain variance estimates.
- Implemented in strs using cuminc option.
- Also available in SEER*Stat.
Using strs

- strs will perform the calculations described in Cronin and Feuer.
- Adding the cuminc option will do this.
- For example,

```
strs using popmort, br(0(1)10) mergeby(year sex age) by(agegrp) ///
cuminc savgroup(cuminc grp,replace)
```

Colon cancer: crude probabilities

```
. use grouped, clear
. list start end a...agegrp == 1, noobs

<table>
<thead>
<tr>
<th>start</th>
<th>end</th>
<th>agegrp</th>
<th>cp</th>
<th>F</th>
<th>cr_e2</th>
<th>ci_dc</th>
<th>ci_do</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>45-59</td>
<td>0.7499</td>
<td>0.2501</td>
<td>0.7558</td>
<td>0.2432</td>
<td>0.0068</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>45-59</td>
<td>0.6317</td>
<td>0.3683</td>
<td>0.6419</td>
<td>0.3559</td>
<td>0.0124</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>45-59</td>
<td>0.5672</td>
<td>0.4328</td>
<td>0.5814</td>
<td>0.4151</td>
<td>0.0177</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>45-59</td>
<td>0.5273</td>
<td>0.4727</td>
<td>0.5457</td>
<td>0.4497</td>
<td>0.0229</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>45-59</td>
<td>0.4943</td>
<td>0.5067</td>
<td>0.5168</td>
<td>0.4775</td>
<td>0.0282</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>45-59</td>
<td>0.4710</td>
<td>0.5290</td>
<td>0.4981</td>
<td>0.4953</td>
<td>0.0336</td>
</tr>
<tr>
<td>6</td>
<td>7</td>
<td>45-59</td>
<td>0.4561</td>
<td>0.5459</td>
<td>0.4892</td>
<td>0.5046</td>
<td>0.0393</td>
</tr>
<tr>
<td>7</td>
<td>8</td>
<td>45-59</td>
<td>0.4398</td>
<td>0.5602</td>
<td>0.4770</td>
<td>0.5150</td>
<td>0.0451</td>
</tr>
<tr>
<td>8</td>
<td>9</td>
<td>45-59</td>
<td>0.4336</td>
<td>0.5664</td>
<td>0.4769</td>
<td>0.5151</td>
<td>0.0513</td>
</tr>
<tr>
<td>9</td>
<td>10</td>
<td>45-59</td>
<td>0.4248</td>
<td>0.5752</td>
<td>0.4744</td>
<td>0.5174</td>
<td>0.0578</td>
</tr>
</tbody>
</table>
```

Colon cancer: crude probabilities

```
. use grouped, clear
. list start end agegrp cp F cr_e2 ci_dc ci_do if agegrp == 3, noobs

<table>
<thead>
<tr>
<th>start</th>
<th>end</th>
<th>agegrp</th>
<th>cp</th>
<th>F</th>
<th>cr_e2</th>
<th>ci_dc</th>
<th>ci_do</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>75+</td>
<td>0.5424</td>
<td>0.4576</td>
<td>0.5994</td>
<td>0.3816</td>
<td>0.0760</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>75+</td>
<td>0.4188</td>
<td>0.5812</td>
<td>0.5104</td>
<td>0.4584</td>
<td>0.1228</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>75+</td>
<td>0.3531</td>
<td>0.6469</td>
<td>0.4772</td>
<td>0.4842</td>
<td>0.1626</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>75+</td>
<td>0.2994</td>
<td>0.7006</td>
<td>0.4527</td>
<td>0.5015</td>
<td>0.1991</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>75+</td>
<td>0.2585</td>
<td>0.7415</td>
<td>0.4413</td>
<td>0.5086</td>
<td>0.2329</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>75+</td>
<td>0.2187</td>
<td>0.7813</td>
<td>0.4253</td>
<td>0.5173</td>
<td>0.2640</td>
</tr>
<tr>
<td>6</td>
<td>7</td>
<td>75+</td>
<td>0.1860</td>
<td>0.8140</td>
<td>0.4162</td>
<td>0.5218</td>
<td>0.2923</td>
</tr>
<tr>
<td>7</td>
<td>8</td>
<td>75+</td>
<td>0.1571</td>
<td>0.8429</td>
<td>0.4092</td>
<td>0.5246</td>
<td>0.3183</td>
</tr>
<tr>
<td>8</td>
<td>9</td>
<td>75+</td>
<td>0.1302</td>
<td>0.8698</td>
<td>0.3997</td>
<td>0.5290</td>
<td>0.3418</td>
</tr>
<tr>
<td>9</td>
<td>10</td>
<td>75+</td>
<td>0.1090</td>
<td>0.8910</td>
<td>0.4004</td>
<td>0.5278</td>
<td>0.3632</td>
</tr>
</tbody>
</table>
Limitation of life table approach

- Only estimated at end of interval.
- Large age groups: The crude probability of death for someone aged 60 will be different to someone aged 70.
- Limited to how many groups you can investigate (no borrowing of strength).
- We will now move to model based estimates of crude probabilities.

Crude probabilities from models

- We can obtain an estimate of relative survival from the various models we have described (Poisson, flexible parametric, cure).
- For individual level models this gives an individual based prediction of the net probability of death due to cancer (1 - relative survival).
- However, we use the ideas from competing risks theory to also calculate the crude probability of death due to cancer and due to other causes.

Brief mathematical details[4]

\[ h^*(t) \] - expected mortality rate
\[ \lambda(t) \] - excess mortality rate
\[ h(t) = h^*(t) + \lambda(t) \] - all-cause mortality rate
\[ S^*(t) \] - expected survival
\[ R(t) \] - relative survival
\[ S(t) = S^*(t)R(t) \] - All cause survival

Net Prob of Death = \[ 1 - R(t) = 1 - \exp \left( - \int_0^t \lambda(u)du \right) \]

Crude Prob of Death (cancer) = \[ \int_0^t S^*(u)R(u)\lambda(u)du \]

Crude Prob of Death (other causes) = \[ \int_0^t S^*(u)R(u)h^*(u)du \]
Estimating the crude probabilities

- We will use flexible parametric relative survival models.
- For any particular covariate pattern, we can estimate the excess mortality rate, $\hat{\lambda}(t|x_i)$.
- The overall survival for individual $i$ is $S_i^*(t)\hat{R}_i(t)$.
- We plug these into the equations in the previous slide to estimate the crude probabilities.
- We have to do the integration numerically.

The stpm2cm command

- The stpm2cm command is a post estimation command.
- It assumes that you have already fitted a relative survival model.
- The prediction is for a particular covariate pattern, specified using the at() option.
- You need to specify the values for variables in the life table that the prediction is for (age, sex, calendar year).
- Note that even if you model age as groups then you still have to specify an exact age, calendar year etc, that the prediction is for.
- For example,

```
stpm2 agegrp2-agegrp4, scale(hazard) bhazard(rate) df(b) tvc(agegrp2-agegrp4) dftvc(b)
```

Crude probabilities (woman aged 75)

![Crude probabilities graph](image-url)
Comparison with net probability

![Graph showing comparison with net probability.

Male Hodgkin lymphoma: trends in 5-year RSR

![Graph showing trends in 5-year RSR for different age groups.

Male Hodgkin lymphoma: trends in 10-year RSR

![Graph showing trends in 10-year RSR for different age groups.

Sandra Eloranta, Paul C. Lambert, Jan Sjöberg, Therese M.L. Andersson, Magnus Björkholm, and Paul W. Dickman

Partitioning of excess mortality in population-based cancer patient survival studies using flexible parametric survival models

Sandra Eloranta1*, Paul C Lambert2, Therese ML Andersson3, Karina Czene1, Per Hall1, Magnus Björkholm3 and Paul W Dickman1

RESEARCH ARTICLE Open Access
Breast cancer survival comparison [5]

- Data from England and Norway.
- The data consists of
  - 303,657 women from England.
  - 24,919 women from Norway.
  - Year of Diagnosis was between 1996 and 2004.
- Extension of

Quantifying differences in breast cancer survival between England and Norway
Paul C. Lambert a,b,*, Lars Holmberg c, Fredrik Sandin d, Freddie Bray e, f, Karen M. Linklater g, Arnie Purushotham h, David Robinson e, Henrik Møller i

**A B S T R A C T**

Background: Survival from breast cancer is lower in the UK than in some other European countries. We compared survival in England and Norway by age and time from diagnosis. Methods: We included 303,648 English and 24,919 Norwegian cases of breast cancer diagnosed 1996–2004 using flexible parametric relative survival models, enabling improved quantification of differences in survival. Crude probabilities were estimated to partition the probability of death due to all causes into that due to cancer and other causes and to estimate the number of “avoidable” deaths. Results: England had lower relative survival for all ages with the difference increasing with age. Much of the difference was due to higher excess mortality in England in the first few months after diagnosis. Older patients had a higher proportion of deaths due to other causes. At 5 years post diagnosis, a woman aged 85 in England had probabilities of 0.35 of dying of cancer and 0.32 of dying of other causes, whilst in Norway they were 0.26 and 0.35. By eight years the number of “avoidable” all-cause deaths in England was 1020 with the number of “avoidable” breast cancer related deaths 1489. Conclusion: Lower breast cancer survival in England is mainly due to higher mortality in the first year after diagnosis. Crude probabilities aid our understanding of the impact of disease on individual patients and help assess different treatment options.

Flexible parametric models

\[ H(t) = H^*(t) + \Lambda(t) \]

- Model \( \ln [\Lambda(t)] \) scale which includes terms for
  - Baseline hazard (time) - Splines (6 parameters)
  - Country - 1 dummy covariate
  - Age - Splines (4 parameters)
  - Age×Country - (4 parameters)
  - Country×Time - Splines (3 parameters)
  - Age×Time - \( 4 \times 3 = 12 \) parameters

- Results extremely robust to number and locations of the knots.
Excess mortality rate ratios

![Graph showing excess mortality rate ratios for different ages and years from diagnosis](image)

Excess mortality rate differences

![Graph showing excess mortality rate differences for different ages and years from diagnosis](image)

Breast cancer crude probabilities, England

![Graph showing breast cancer crude probabilities for different ages and years from diagnosis](image)
Breast cancer crude probabilities, Norway

![Graph showing breast cancer probabilities for different ages and years from diagnosis.]

Probabilities at 5 Years

<table>
<thead>
<tr>
<th>Age</th>
<th>Crude Probabilities</th>
<th>Net Probabilities</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cancer</td>
<td>Other</td>
</tr>
</tbody>
</table>
| England  
45  | 0.15   | 0.01  | 0.84  | 0.15 | 0.85  |
| 55  | 0.13   | 0.02  | 0.85  | 0.13 | 0.87  |
| 65  | 0.16   | 0.06  | 0.78  | 0.17 | 0.83  |
| 75  | 0.25   | 0.14  | 0.61  | 0.27 | 0.73  |
| 85  | 0.35   | 0.32  | 0.33  | 0.42 | 0.58  |
| Norway  
45  | 0.13   | 0.01  | 0.86  | 0.13 | 0.87  |
| 55  | 0.11   | 0.02  | 0.87  | 0.11 | 0.89  |
| 65  | 0.13   | 0.05  | 0.82  | 0.13 | 0.87  |
| 75  | 0.19   | 0.12  | 0.68  | 0.21 | 0.79  |
| 85  | 0.26   | 0.35  | 0.39  | 0.31 | 0.69  |

References


