Cancer Survival in Florida 1999-2003

or Why Rates are Harder Than Counts

Anders Alexandersson

Florida Cancer Data System

Table of contents

1. Introduction

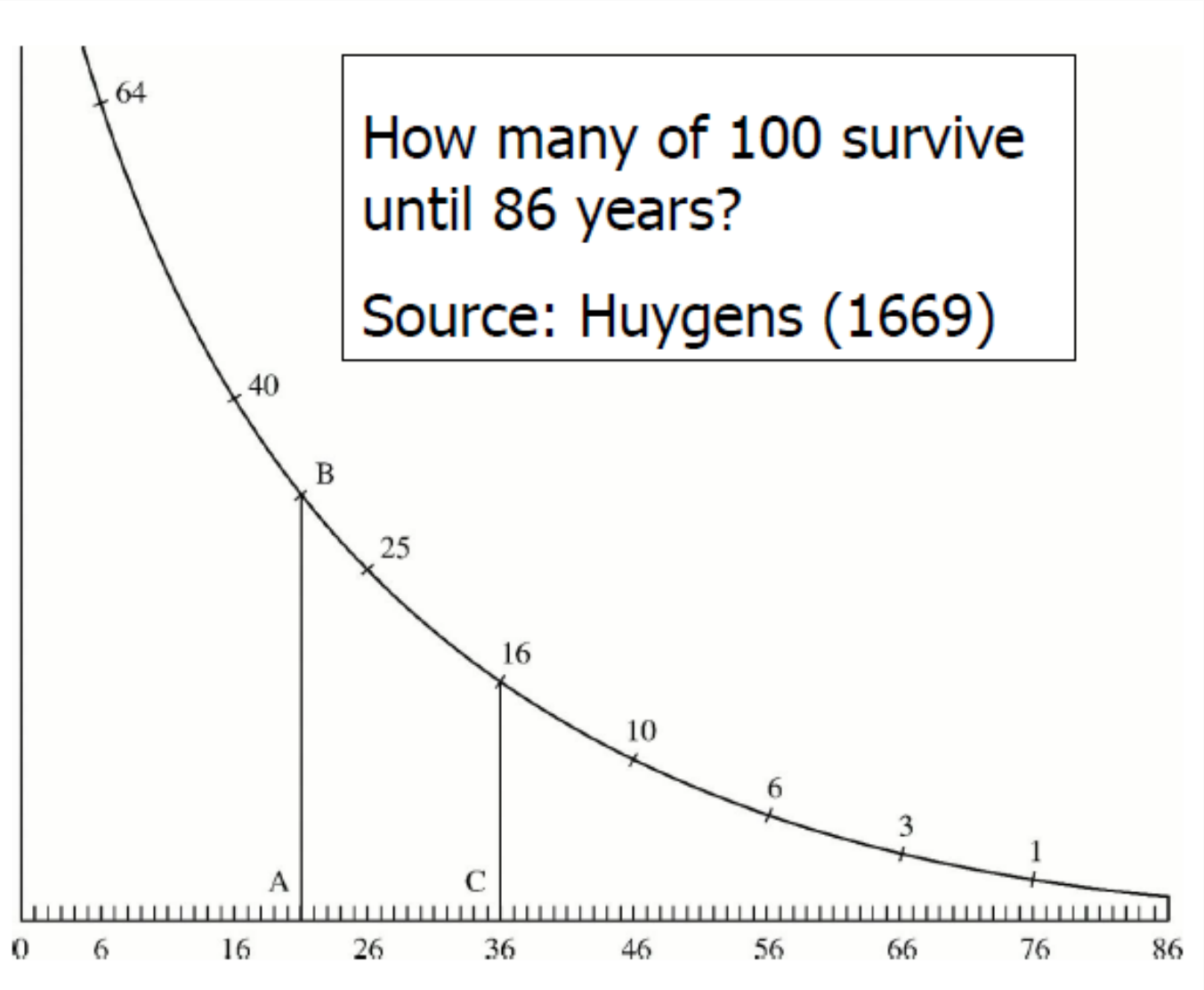
2. Data Management

3. Survival analysis

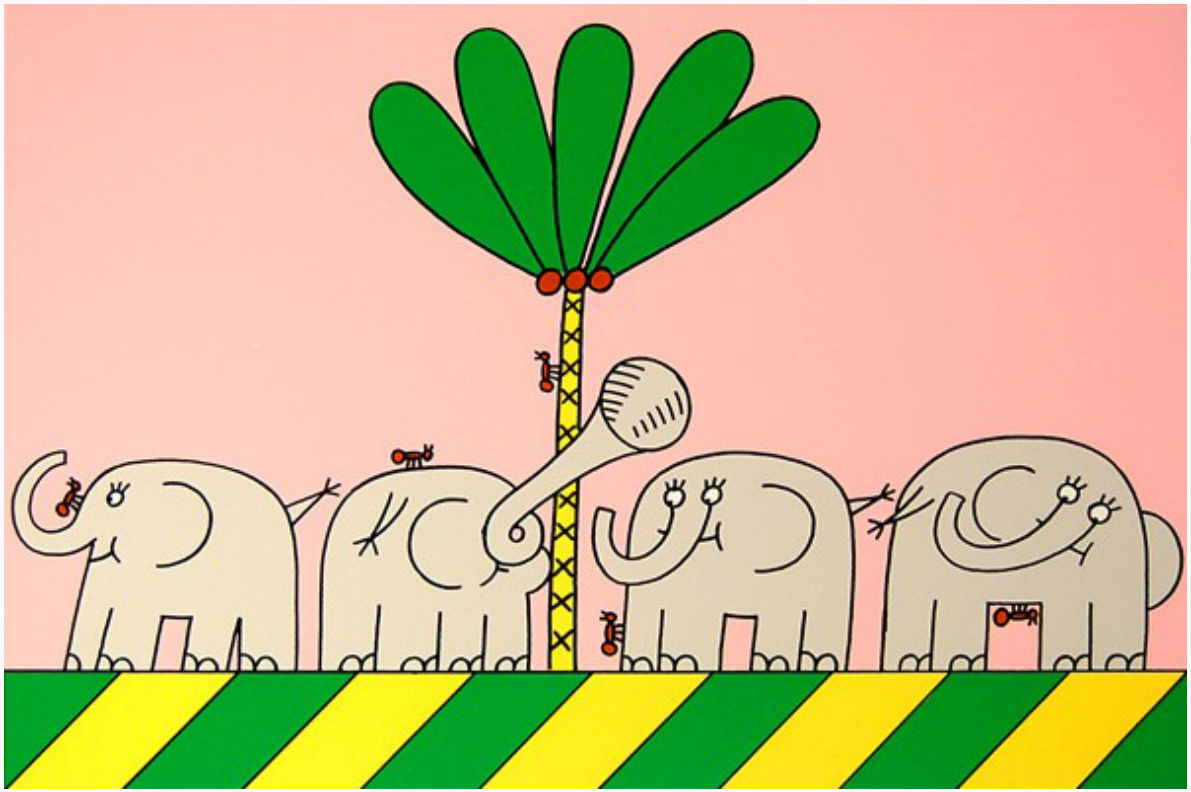
4. Reporting

Introduction

Early survival analysis used counts

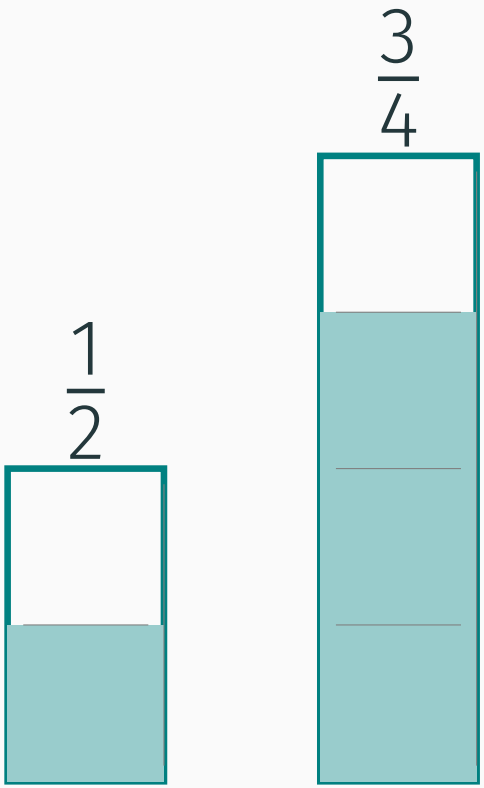


Counts are easy



”5 ants are more than 4 elephants”, Swedish children TV.

Rates are hard



Average ratio = 62.5%

Ratio of averages = 67%

The order of calculation produces different results.

Death rates are hard

G. Udny Yule
(1934)

Death (Mortality) rates can be either an average ratio or a ratio of averages.

Survival rates are hard

Pohar Perme et al.
(2012)

Net survival is the average ratio of overall and population survival.

Relative survival ratio is the ratio of averages of overall and population survival.

The FCDS 2017 Monograph uses net survival

Survival analysis is time-to-event analysis.

The focus is on net survival in a relative framework.

It is reproducible research using Stata.

Data science is used for reproducibility

Data science = data management + statistical analysis + programming

Data Management

Import data

import delimited

Tidy data

Check for unique identifiers

1. **isid** Patient_Id_Number_N20 Sequence_Number_Central_N380
2. Run the SAS program ”CalculateSurvivalTimeInMonths.sas”
3. Rename variables to your liking
4. **isid** pid_20 record_order

Transform data

Create the analysis dataset

1. Create new variables such as FCDS site group, ICSS weights
2. Specify exclusions, e.g., omit children
3. Review the data
4. Create population mortality file from Human Mortality Database

Survival analysis

Survival analysis concepts

Analysis time t is time at risk:

$$t = \frac{time - origin}{scale}$$

Examples are years or months since diagnosis of cancer.

Survival time T is the time until a failure event. Examples are years or months since diagnosis of cancer until death.

The **survival function** $S(t)$ is the probability of surviving beyond t, the probability that the survival time T is larger than the specific time t:

$$S(t) = P(T > t)$$

Survival analysis measures

Net survival		
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 survival		
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

Survival analysis frameworks

Crude and net survival distinguish between two causes of death: death due to cancer and death due to other causes.

Overall (a.k.a. observed or all-cause) survival, and relative survival ratio do NOT make this distinction.

Crude and net survival can be estimated in the cause-specific or relative framework.

Survival analysis approaches and recommended FCDS usage

Measure	Framework: Cause-specific	Framework: Relative
Crude	Registry-based randomized controlled trial (RRCT)	Risk communication
Net	Causality with observational data	Life tables

Pohar Perme estimates

Net survival in the relative framework using life tables

Cancer registries often prefer net survival (ignores competing risks) in a relative framework (does not require cause of death).

The **Pohar Perme estimator** corrects for deaths due to other causes.

FCDS has *continuous survival time* but birth month and birth day are not releasable. This suggests Pohar Perme estimation using *life tables*.

The user-written Stata command stnet[3] is designed for this.

stnet command

1) Declare data to be survival-time data:

```
. use doh if site_10group==1, clear
. stset surv_year, failure(vital_1760==0) id(pid_20)
```

2) Look at the output. Fix any problems.

3) Use the **stnet** command:

```
stnet using popmort9913 ///
    if inrange(dx_year,1999,2003) [iw=icss1], ///
    mergeby(_year sex _age) breaks(0(0.083333333)10)  ///
    diagdate(date_dx) birthdate(dob) standstrata(agegr)  ///
    savstand(agestand__NS1, replace)
```

stnet output

```
. use agestand__NS1, clear
. list if inlist(end,1,5,10), noobs
```

start	end	cns	secns	locns	upcns
.9167	1	0.5047	0.0024	0.5000	0.5093
4.917	5	0.2090	0.0021	0.2048	0.2132
9.917	10	0.1534	0.0024	0.1488	0.1581

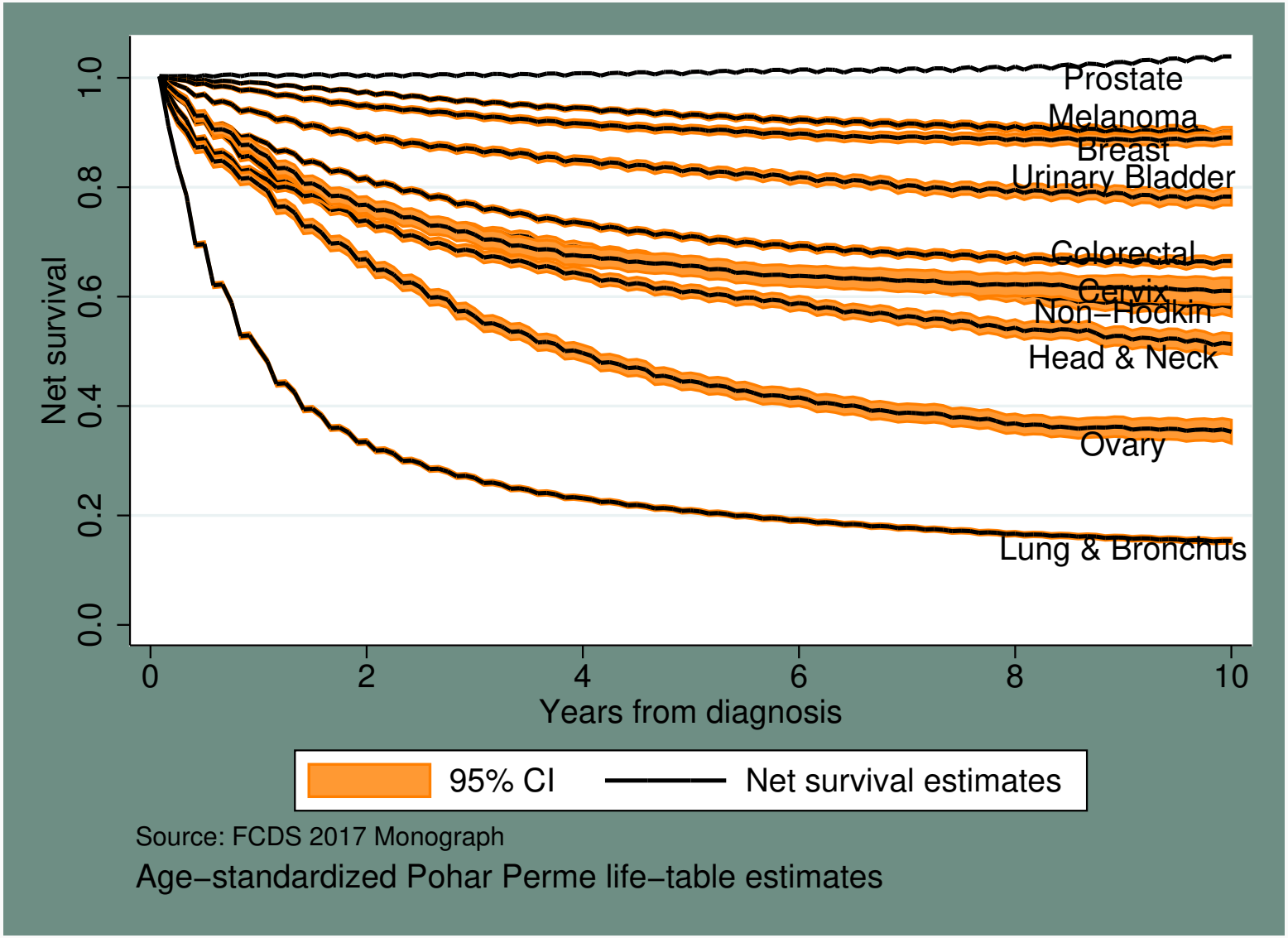
Reporting

Reporting tools

texdoc creates dynamic reports in LaTeX.

tabout creates publication-quality summary tables.

Net survival by cancer site, Dx 1999-2003

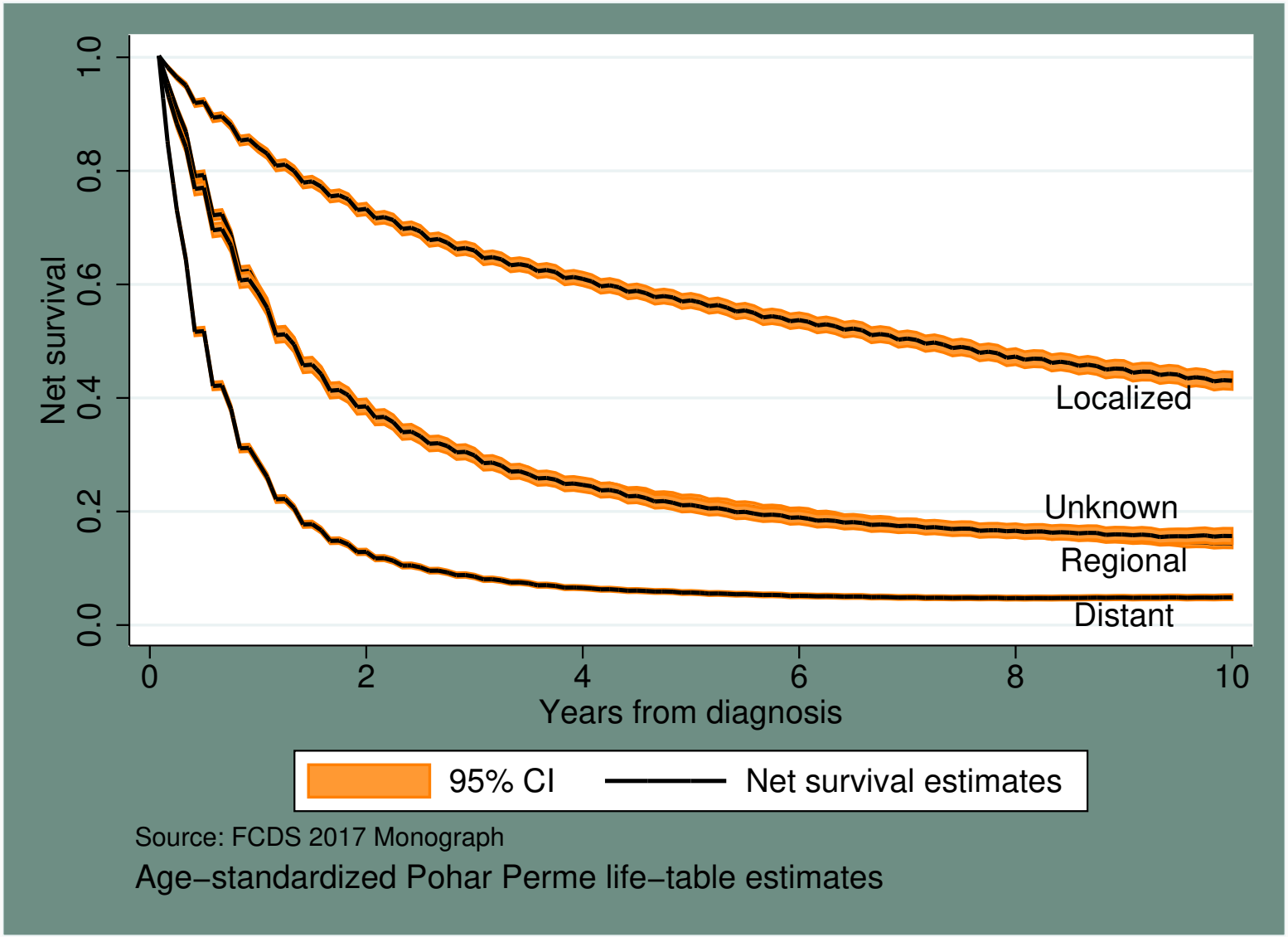


Net survival (%) by cancer site, Dx 1999-2003

Cancer	1-year	5-year	10-year
Lung & Bronchus	50.5	20.9	15.3
Prostate	100.6	101.1	103.9
...	...	...	...
Cervix	86.4	65.7	61.1

Source: FCDS 2017 Monograph[1]

Net survival of lung cancer by stage, Dx 1999-2003



Source: FCDS 2017 Monograph

Age–standardized Pohar Perme life–table estimates

Net survival (%) of lung cancer, Dx 1999-2003

	1-year	5-year	10-year
Age Standardized			
Total	50.5	20.9	15.3
Sex			
Male	46.6	18.0	13.2
Female	55.2	24.4	18.0
Race			
...	...	...	...

Source: FCDS 2017 Monograph[1]

Read the FCDS 2017 Monograph[1] and the Technical Report[2]. ☺

References

[1] A. Alexandersson. Cancer survival in florida 1999-2003 with 10-year follow-up. FCDS Monograph, 2017.

[2] A. Alexandersson. Survival analysis of the florida cancer data system: A data science project using stata. FCDS Technical Report, 2017.

[3] E. Coviello, P. Dickman, K. Seppä, and A. Pokhrel. Estimating net survival using a life-table approach. *Stata Journal*, 15(1):173–185, 2015.