Observations on the Bills of Mortality: The first statistical analysis

Unit 1 Lecture 2

Jonathan Auerbach
STAT 489 Pre-Cap Prof Development
jauerba@gmu.edu

September 14, 2021

Learning Objectives

After this lecture, you will be able to:

1. Describe the statistical analysis conducted by John Graunt.
2. Calculate period life expectancy at birth from a life table.
3. Explain in what sense life expectancy is a fair representation of a population's longevity.
4. Graph a simple tree diagram using `ggtree`. See Appendix for `R` code.

These slides use the following `R` packages

Setup:

```r
library("tidyverse")
library("treeio")
library("ggtree")
library("knitr")
theme_set(theme_bw())
```

The package ggtree is not available on the Comprehensive R Archive Network (CRAN). Install it from Bioconductor:

```r
install.packages("BiocManager")
BiocManager::install("ggtree")
```

Observations on the Bills of Mortality

- The bills reported the number of burials (deaths) in London.
  - Sporadic publication started in the sixteenth century. Weekly publication began in 1603—Londoners could subscribe for a fee.
  - Bills counted deaths by cause, e.g. plague, measles, and old age.
  - Londoners used the bills as a plague warning system: to identify outbreaks and determine when to leave or return to the city.

- Graunt's *Observations on the Bills of Mortality* (1662) was the first publication to analyze the bills statistically and answer the most pressing demographic questions of the time.
  - Among the 106 observations listed in the book's index, he found London's population was lower than previously estimated, and the population lost after a plague outbreak rebounded faster.

- To answer these questions, Graunt calculated several new statistics.
  - The most famous and our focus: (period) life expectancy at birth.
  - More importantly, Graunt's analysis demonstrated the value of statistics. Cities raced to collect more data, initiating the field.

Bill of Mortality (Company of Parish Clerks, 1665)

A generall Bill for this present year, ending the 19 of *December* 1665. according to the Report made to the KINGS most Excellent Majesty.

By the Company of Parish Clerks of *London*, &c.

	Buried	Plag.		Buried	Plag.		Buried	Plag.		Buried	Plag.
St Albans Woodstreet	200	121	St Clements Eastcheap	38	20	St Margaret Moses	38	25	St Michael Cornhill	104	52
St Alhallowes Barking	514	330	St Dionis Back-church	78	27	St Margaret Newfishst	114	66	St Michael Crookedla.	179	133
St Alhallowes Breadst	35	16	St Dunstans East	265	150	St Margaret Pattons	49	24	St Michael Queenhit	203	122
St Alhallowes Great	455	426	St Edmunds Lumbard.	70	36	St Mary Abchurch	99	54	St Michael Quern	44	18
St Alhallowes Honila	10	5	St Ethelborough	195	106	St Mary Aldermanbury	181	109	St Michael Royall	152	116
St Alhallowes Lesse	239	175	St Faiths	104	70	St Mary Aldermary	105	75	St Michael Woodstreet	122	62
St Alhall. Lumbardstr.	90	62	St Fosters	144	105	St Mary le Bow	64	36	St Mildred Breadstreet	59	26
St Alhallowes Staining	185	112	St Gabriel Fen-church	69	39	St Mary Bothaw	55	30	St Mildred Poultrey	68	46
St Alhallowes the Wall	500	356	St George Botolphlane	41	27	St Mary Colechurch	17	6	St Nicholas Acons	46	28
St Alphage	271	115	St Gregories by Pauls	376	232	St Mary Hill	94	64	St Nicholas Colcabby	125	91
St Andrew Hubbard	71	25	St Hellens	108	75	St Mary Mounthaw	56	37	St Nicholas Olaves	90	62
St Andrew Vndershaft	274	189	St James Dukes place	262	190	St Mary Summerset	342	262	St Olaves Hartstreet	237	160
St Andrew Wardrobe	476	308	St James Garlickhithe	189	118	St Mary Stayninge	47	27	St Olaves Iewry	54	32
St Anne Aldersgate	282	197	St John Baptist	138	83	St Mary Woolchurch	65	33	St Olaves Siluerstreete	250	132
St Anne Blacke-Friers	652	467	St John Euangelist	9		St Mary Woolnoth	75	38	St Pancras Soperlane	30	15
St Antholins Parish	58	33	St John Zacharie	85	54	St Martins Iremonger.	21	11	St Peters Cheape	61	35
St Austins Parish	43	20	St Katherine Coleman	299	213	St Martins Ludgate	196	128	St Peters Cornhill	136	76
St Barthol. Exchange	73	51	St Katherine Creechu.	335	231	St Martins Orgars	110	71	St Peters Pauls Wharfe	114	86
St Bennet Fynch	47	22	St Lawrence Iewry	94	48	St Martins Outwitch	60	34	St Peters Poore	79	47
St Benn. Grace-church	57	41	St Lawrence Pountney	214	140	St Martins Vintrey	417	349	St Stevens Colmanstr.	560	391
St Bennet Pauls Wharf	355	172	St Leonard Eastcheap	42	27	St Matthew Fridaystr.	24	6	St Stevens Walbrooke	34	17
St Bennet Sherehog	11	1	St Leonard Fosterlane	335	255	St Maudlins Milkstreet	44	22	St Swithins	93	56
St Botolph Billingsgate	83	50	St Magnus Parish	103	60	St Maudlins Oldfishstr.	176	121	St Thomas Apostle	163	110
Christs Church	653	467	St Margaret Lothbury	100	66	St Michael Bassishaw	253	164	Trinitie Parish	115	79
St Christophers	60	47									

Buried in the 97 Parishes within the walls, — 15207 *Whereof, of the Plague* — 9887

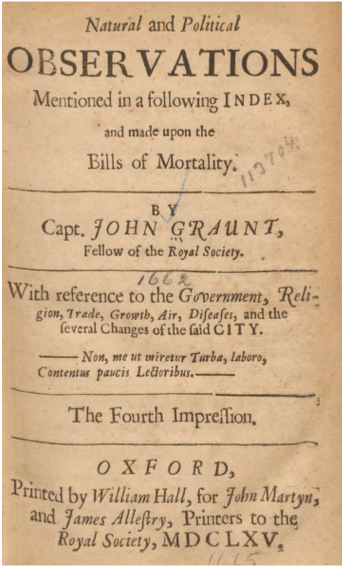
St Andrew Holborne	3958	3103	Bridewell Precinct	230	179	St Dunstans West	958	665	St Saviours Southwark	4235	3446
St Bartholmew Great	493	344	St Botolph Aldersga.	997	755	St George Southwark	1613	1260	St Sepulchres Parish	4509	3746
St Bartholmew Lesse	193	139	St Botolph Algate	4926	4051	St Giles Cripplegate	8069	4838	St Thomas Southwark	475	371
St Bridget	2111	1427	St Botolph Bishopsg.	3464	2500	St Olaves Southwark	4793	2785	Trinity Minories	168	123
									At the Pesthouse	159	156

Buried in the 16 Parishes without the Walls — 41351 *Whereof, of the Plague* — 28888

Source: https://commons.wikimedia.org/wiki/File:Bill_of_Mortality.jpg

John Graunt (1623-1687) and *Observations* (1662)

Captain John Graunt



Natural and *Political*
OBSERVATIONS
Mentioned in a following INDEX,
and made upon the
Bills of Mortality.

BY
Capt. *JOHN GRAUNT*,
Fellow of the *Royal Society*.

With reference to the *Government*, *Religion*, *Trade*, *Growth*, *Air*, *Diseases*, and the ſeveral Changes of the ſaid CITY.

——— *Non, me ut miretur Turba, laboro,*
Contentus paucis Lectoribus. ———

The Fourth Impreſſion.

OXFORD,
Printed by *William Hall*, for *John Martyn*, and *James Alleſtry*, Printers to the *Royal Society*, MDCLXV.

Source: https://en.wikipedia.org/wiki/John_Graunt#/media/File:JohnGraunt.png
http://resource.nlm.nih.gov/2356017R

How did Graunt calculate life expectancy at birth?

- He constructed a life table: the proportion of deaths at each age.
  - Today the proportion is interpreted as the probability a person randomly chosen at birth will die at that age.
    - He grouped ages into stages (0-6, 7-16, 17-26, ..., 67-76, 76-79, 80)
    - Denote the proportion $p_n = \mathbb{P}(\text{“die in stage n”})$
  - Life expectancy at birth is average stage/age attained: $\sum_n n p_n$

- The challenge was that the bills did not record age at death.
  - Graunt only observed the number of deaths from each cause—as well as other records like the number of christenings and weddings.
  - By comparing christenings with causes primarily affecting children, he calculated the death rate among the 0-6 age group to be 9/25.

- Graunt then assumed the death rate was the same at every stage.
  - This is equivalent to modeling survival as a multistage coin-flipping experiment. One survives to stage n by flipping tails “n” times.
  - Coin weight $p = \mathbb{P}(\text{“die in stage n”} \mid \text{“alive in stage n-1”}) \approx 9/25$
  - We will see in a moment that $p_n = (1-p)^{n-1}\, p$.

Graunt's life table in *Observations* (1662)

ving ſeven *Decads* between ſix and 76, we ſought ſix mean proportional numbers between 64, the remainder, living at ſix years, and the one, which ſurvives 76, and find, that the numbers following are practically near enough to the truth; for men do not die in exact proportion, nor in Fractions, from whence ariſes this Table following.

Viz. Of an hundred there dies within the firſt ſix years	36	The third *Decad*	9
The nex ten years, or *Decad*	24	The fourth	6
The 2d *Decad*	15	The next	4
		The next	3
		The next	2
		The next	1

10. From whence it follows, that of the ſaid 100 conceived there remain alive at ſix years end 64.

At

At 16 years end	40	At fifty ſix	6
At twenty ſix	25	At ſixty	3
At thirty ſix	16	At ſeventy ſix	1
At fourty ſix	10	At eighty	0

11. It follows alſo, That of all which have been conceived, there are now alive 40 *per Cent.* above ſixteen years old, 25 above twenty ſix years old, & *ſic deinceps*, as in the above-Table. There are therefore of Aged between 16 and 56 the number of 40, leſs by ſix, *viz.* 34; of between 26 and 66 the number of 25, leſs by three, *viz.* 22: & *ſic deinceps*.

Wherefore ſuppoſing there be 199112 *Males*, and the number between 16 and 56 being 34; it follows there are 34 *per Cent.* of all thoſe *Males* fighting men in *London*, that is 67694, *viz.* near 70000; the

Source: http://resource.nlm.nih.gov/2356017R

Graunt's life table (per hundred births)

```
life_table <-
  tibble(Age       = c(  0, 6,16,26,36,46,56,66,76,80),
         Deaths    = c(  0,36,24,15, 9, 6, 4, 3, 2, 1),
         Survivors = c(100,64,40,25,16,10, 6, 3, 1, 0))
kable(life_table)
```

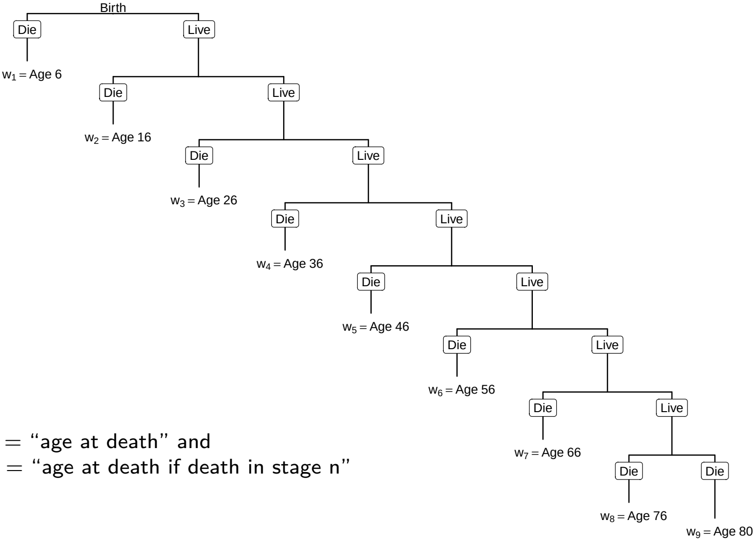
Age	Deaths	Survivors
0	0	100
6	36	64
16	24	40
26	15	25
36	9	16
46	6	10
56	4	6
66	3	3
76	2	1
80	1	0

$\mathbb{P}$(**"die in stage n" | "alive in stage n - 1"**) $\approx 9/25$

```
life_table %>%
  mutate(`Stage` = replace(row_number() - 2, 1, NA),
         `Deaths (approx)` = 100 * (1-9/25)^Stage * 9/25,
         `Deaths (approx)` = 100 * dgeom(Stage, 9/25)) %>%
  kable(digits = 1)
```

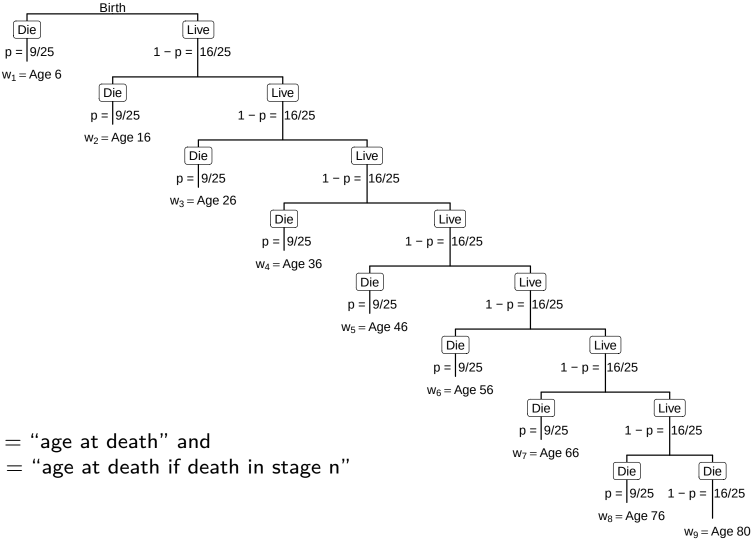
Age	Deaths	Survivors	Stage	Deaths (approx)
0	0	100	NA	NA
6	36	64	0	36.0
16	24	40	1	23.0
26	15	25	2	14.7
36	9	16	3	9.4
46	6	10	4	6.0
56	4	6	5	3.9
66	3	3	6	2.5
76	2	1	7	1.6
80	1	0	8	1.0

Step 1: Enumerate all possible outcomes



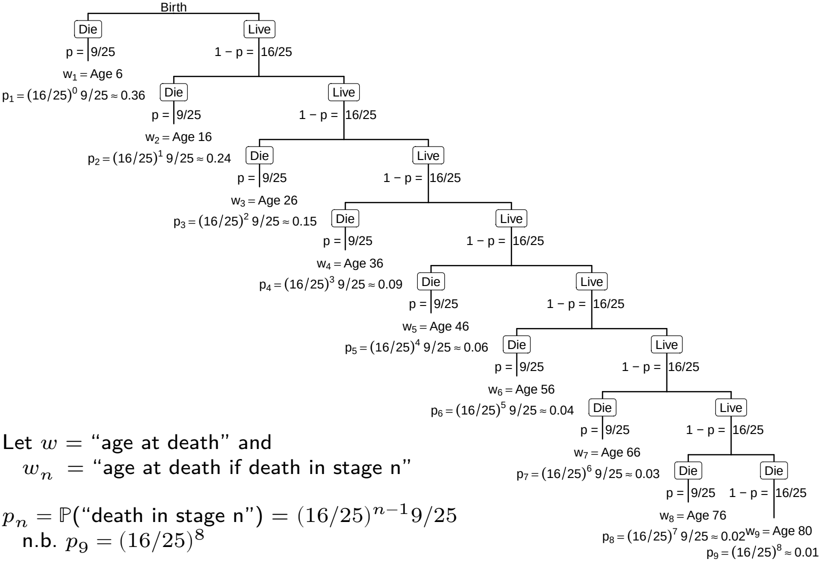
Let w = "age at death" and
w_n = "age at death if death in stage n"

Step 2: Label the probability of outcomes by stage



Let w = "age at death" and
w_n = "age at death if death in stage n"

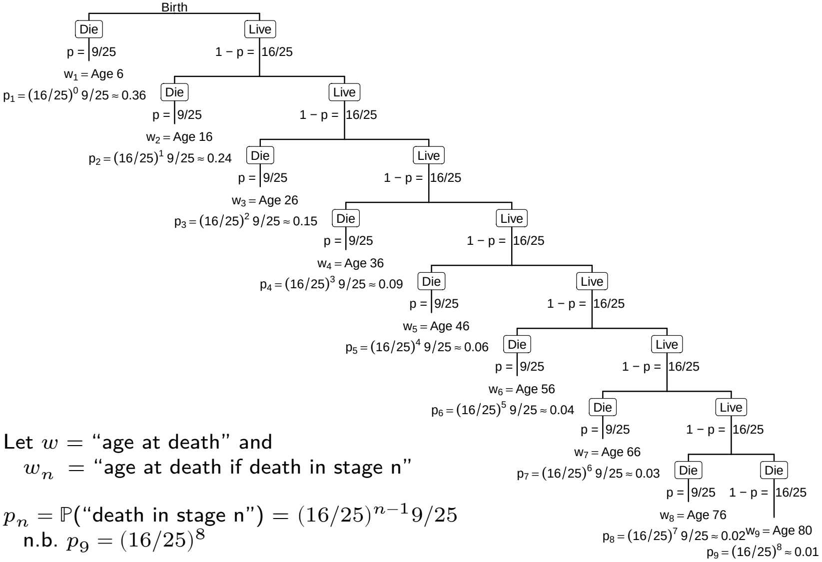
Step 3: Multiply vertical probabilities



Let w = "age at death" and
w_n = "age at death if death in stage n"

$p_n = \mathbb{P}(\text{"death in stage n"}) = (16/25)^{n-1} 9/25$
n.b. $p_9 = (16/25)^8$

Step 4: Add probability-weighted outcomes



Let w = "age at death" and
w_n = "age at death if death in stage n"

$p_n = \mathbb{P}(\text{"death in stage n"}) = (16/25)^{n-1} 9/25$
n.b. $p_9 = (16/25)^8$

$\mathbb{E}[w] = \sum_n w_n p_n \approx 18$

Life expectancy at birth using Graunt's table

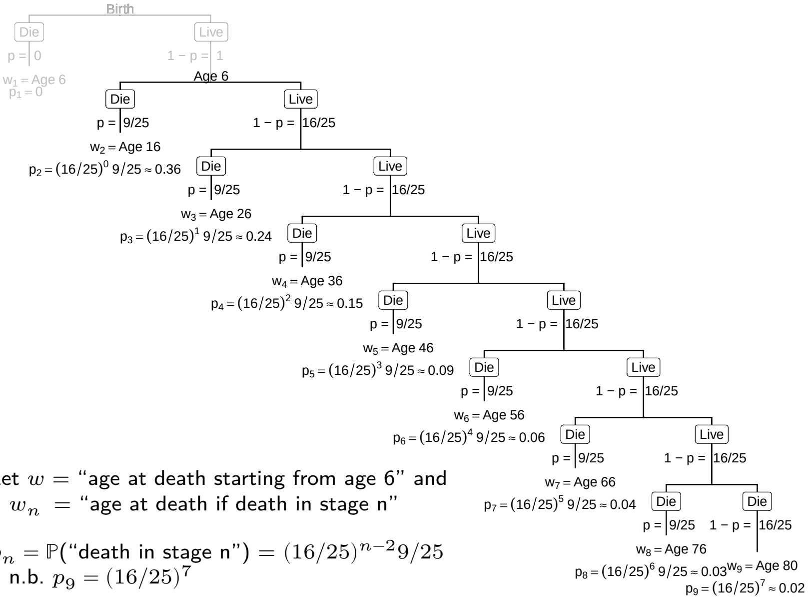
```
life_table %>%
  mutate(`Mid Period Age` = Age - c(0, diff(Age))/2) %>%
  slice_head(n = 3) %>% kable()
```

Age	Deaths	Survivors	Mid Period Age
0	0	100	0
6	36	64	3
16	24	40	11

```
life_table %>%
  mutate(`Mid Period Age` = Age - c(0, diff(Age))/2) %>%
  summarize(`Life Expectancy from Birth` =
    sum(`Mid Period Age` * `Deaths`) / 100) %>% kable()
```

Life Expectancy from Birth
18.19

Life expectancy from age 6 is an interrupted game



Let w = "age at death starting from age 6" and
w_n = "age at death if death in stage n"

$p_n = \mathbb{P}(\text{"death in stage n"}) = (16/25)^{n-2} 9/25$
n.b. $p_9 = (16/25)^7$

$\mathbb{E}[w] = \sum_n w_n p_n \approx 27$

Graunt's life table if starting from age 6

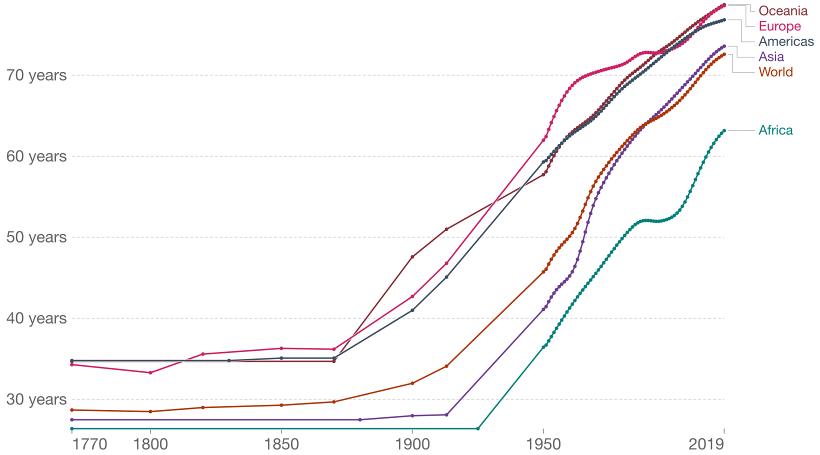
```
life_table %>%
  mutate(`Mid Period Age` = Age - c(0, diff(Age))/2,
         `Deaths Starting from Age 6` =
           replace(100 * Deaths/Survivors[2], 2, 0)) %>%
  kable(digits = 1)
```

Age	Deaths	Survivors	Mid Period Age	Deaths Starting from Age 6
0	0	100	0	0.0
6	36	64	3	0.0
16	24	40	11	37.5
26	15	25	21	23.4
36	9	16	31	14.1
46	6	10	41	9.4
56	4	6	51	6.2
66	3	3	61	4.7
76	2	1	71	3.1
80	1	0	78	1.6

Graunt's methods immediately and widely adopted.

- His analyses were revolutionary and brought him instant fame.
  - Graunt held a number of political offices before publication—already a great achievement given his modest background.
  - But upon completing *Observations*, he was admitted into the Royal Society, the new and elite academic circle of the day.
- Much of his legacy due to his careful assessment of data quality.
  - For example, Graunt thought deaths were systematically misclassified. Plague deaths by as much as 25% during outbreaks.
    - Data collectors were likely bribed to misclassify plague deaths to avoid quarantine policies.
    - Families may also have bribed them to misclassify embarrassing diseases like syphilis.
- Today, (period) life expectancy at birth is the most common measure of population health.
  - World life expectancy has doubled over the past century, although substantial inequality exists among countries.

Life expectancy (1770-2019, Our World in Data)



Source: https://ourworldindata.org/life-expectancy

References

1. Hacking, Ian. The emergence of probability: A philosophical study of early ideas about probability, induction and statistical inference. Cambridge University Press, 2006.

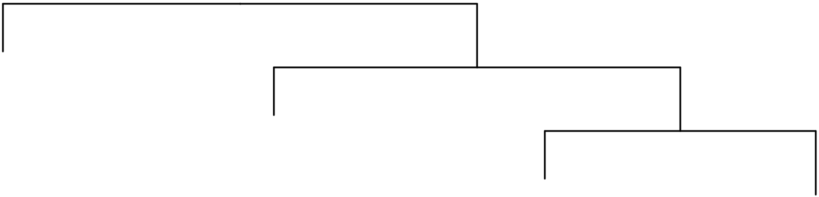
2. Hald, Anders. A history of probability and statistics and their applications before 1750. John Wiley & Sons, 2005.

3. Roser, Max, Esteban Ortiz-Ospina, and Hannah Ritchie. Life expectancy. Our World in Data. 2021.

4. Sutherland, Ian. John Graunt: a tercentenary tribute. Journal of the Royal Statistical Society. 1963.

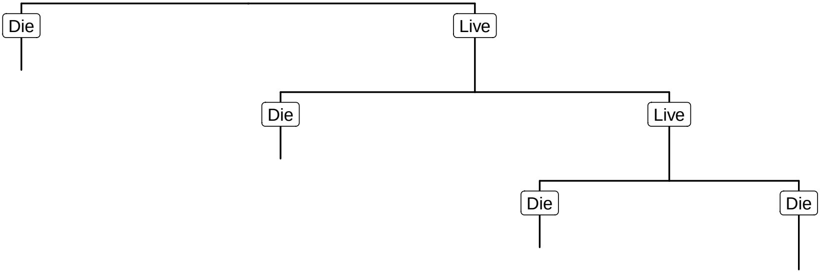
Appendix: Newick Representation of Decision Tree

```
# Tree coded using Newick format
## parens. denote grouping of terminal nodes
## c.f. https://en.wikipedia.org/wiki/Newick_format

tree_text <- "(b:1.5,(c:1.5,(d:1.5, e:5)))a;"
tree_data <- treeio::read.newick(text = tree_text)
tree_data$edge.length[c(2, 4, 6)] <- 2
tree_labels <- tibble(label = letters[1:5],
  outcome = paste0("Age~", seq(-4,36,10)),
  probability = paste0(paste0("p[",0:4), "] == ",
    paste0("(16/25)^{",-1:3,"}", c(rep("~9/25", 4), ""))))

ggtree(tree_data) + layout_dendrogram()
```



Appendix: Graph a lightly annotated tree

```
(decision_tree_unlabeled <-
ggtree(tree_data) %<+% tree_labels +
  theme(plot.margin = unit(c(0,0,10,10), "mm")) +
  layout_dendrogram() + annotate("label",
     x = -sort(rep(seq(1, 5, 2), 2), decreasing = TRUE)-.5,
     y = c(1, 2.75, 2, 3.5, 3, 4),
     label = c(rep(c("Die", "Live"), 2), "Die", "Die")))
```



Appendix: Graph an annotated tree

```
(decision_tree_labeled <-
   decision_tree_unlabeled +
   annotate("text", x = -6.25, y = 2, label = "Birth") +
   geom_tiplab(geom = "text", aes(label = outcome),
               parse = TRUE, vjust = 1, hjust = .5) +
   geom_tiplab(geom = "text", aes(label = probability),
               parse = TRUE, vjust = 2.5, hjust = .6))
```