Naivni Bayesov klasifikator in nomogrami

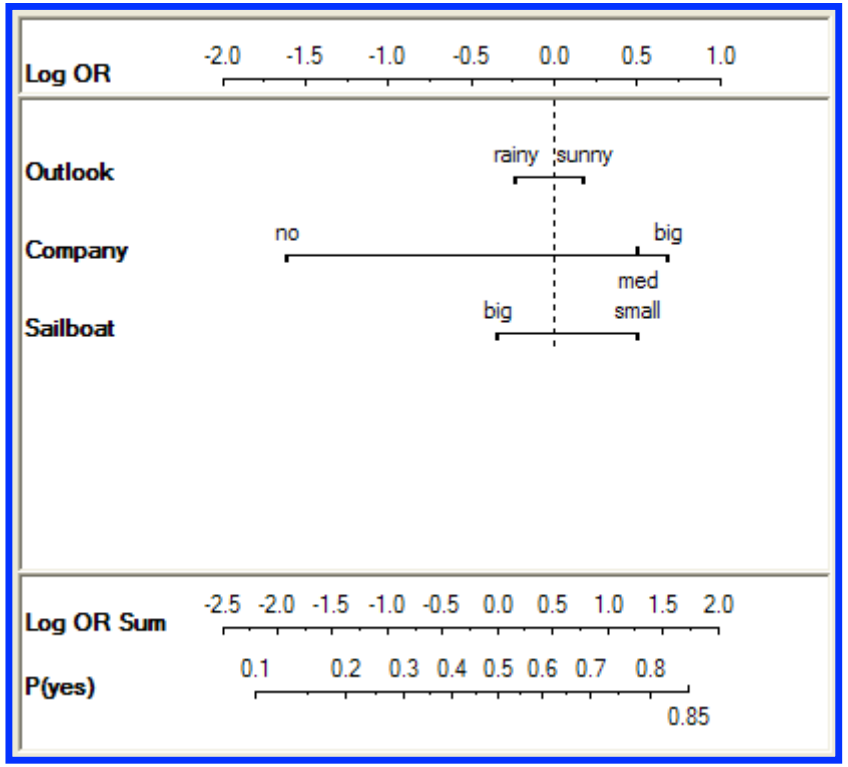
Primer

#	Outlook	Company	Sailboat	Sail
1	rainy	big	big	yes
2	rainy	big	small	yes
3	rainy	med	big	yes
4	rainy	med	small	yes
5	sunny	big	big	yes
6	sunny	big	small	yes
7	sunny	med	big	yes
8	sunny	med	big	yes
9	sunny	med	small	yes
10	sunny	no	small	yes
11	sunny	no	big	no
12	rainy	med	big	no
13	rainy	no	big	no
14	rainy	no	big	no
15	rainy	no	small	no
16	rainy	no	small	no
17	sunny	big	big	no
18	sunny	big	small	no
19	sunny	med	big	no
20	sunny	med	big	no

#	Outlook	Company	Sailboat	Sail
A	sunny	big	big	?
B	rainy	big	small	?
C	sunny		small	?
D	rainy			?

#	Outlook	Company	Sailboat	Sail
1	rainy	big	big	yes
2	rainy	big	small	yes
3	rainy	med	big	yes
4	rainy	med	small	yes
5	sunny	big	big	yes
6	sunny	big	small	yes
7	sunny	med	big	yes
8	sunny	med	big	yes
9	sunny	med	small	yes
10	sunny	no	small	yes
11	sunny	no	big	no
12	rainy	med	big	no
13	rainy	no	big	no
14	rainy	no	big	no
15	rainy	no	small	no
16	rainy	no	small	no
17	sunny	big	big	no
18	sunny	big	small	no
19	sunny	med	big	no
20	sunny	med	big	no

#	Outlook	Company	Sailboat	Sail
A	sunny	big	big	? yes (0.63)
B	rainy	big	small	? yes (0.73)
C	sunny		small	? yes (0.67)
D	rainy			? yes (0.44)



Naivni Bayesov klasifikator

verjetno dogodka e pri pogoju X

X je množica atribut-vrednost

$$p(e|X) = p(e)\frac{p(X|e)}{p(X)}$$

vrednost atributa i

$$p(e|X) = p(e)\frac{\prod p(x_i|e)}{p(X)}$$

pogojna verjetnost

razmerje verjetnost dogodka e (odds for event e)

$$\frac{p(e|X)}{p(\bar{e}|X)} = \frac{p(e)\prod p(x_i|e)}{p(\bar{e})\prod p(x_i|\bar{e})}$$

$$\log\frac{p(e|X)}{p(\bar{e}|X)} = \log\frac{p(e)}{p(\bar{e})} + \sum_i \log\frac{p(x_i|e)}{p(x_i|\bar{e})}$$

Nomogram

$$\log \frac{p(e|X)}{p(\bar{e}|X)} = \log \frac{p(e)}{p(\bar{e})} + \sum_i \log \frac{p(x_i|e)}{p(x_i|\bar{e})}$$

$$\log \frac{p}{\bar{p}} = f$$

$$\frac{p}{\bar{p}} = \frac{p}{1-p} = \mathrm{e}^f$$

$$p = \frac{1}{1+\mathrm{e}^{-f}}$$

Nomogram

$$f = \log \frac{p(e)}{p(\bar{e})} + \sum_i \log \frac{p(x_i|e)}{p(x_i|\bar{e})}$$

$$p = \frac{1}{1 + \mathrm{e}^{-f}}$$

$$p(e) = \frac{N_e}{N}$$

$$p(x_i|e) = \frac{N_{x_i,e}}{N_e}$$

#	Outlook	Company	Sailboat	Sail
1	rainy	big	big	yes
2	rainy	big	small	yes
3	rainy	med	big	yes
4	rainy	med	small	yes
5	sunny	big	big	yes
6	sunny	big	small	yes
7	sunny	med	big	yes
8	sunny	med	big	yes
9	sunny	med	small	yes
10	sunny	no	small	yes
11	sunny	no	big	no
12	rainy	med	big	no
13	rainy	no	big	no
14	rainy	no	big	no
15	rainy	no	small	no
16	rainy	no	small	no
17	sunny	big	big	no
18	sunny	big	small	no
19	sunny	med	big	no
20	sunny	med	big	no

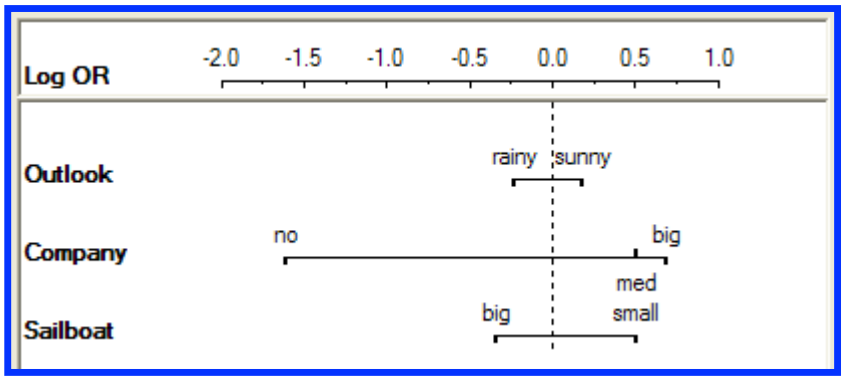
$$f = \log \frac{p(e)}{p(\bar{e})} + \sum_i \log \frac{p(x_i|e)}{p(x_i|\bar{e})}$$

$$p = \frac{1}{1 + \mathrm{e}^{-f}}$$

$$p(e) = \frac{N_e}{N}$$

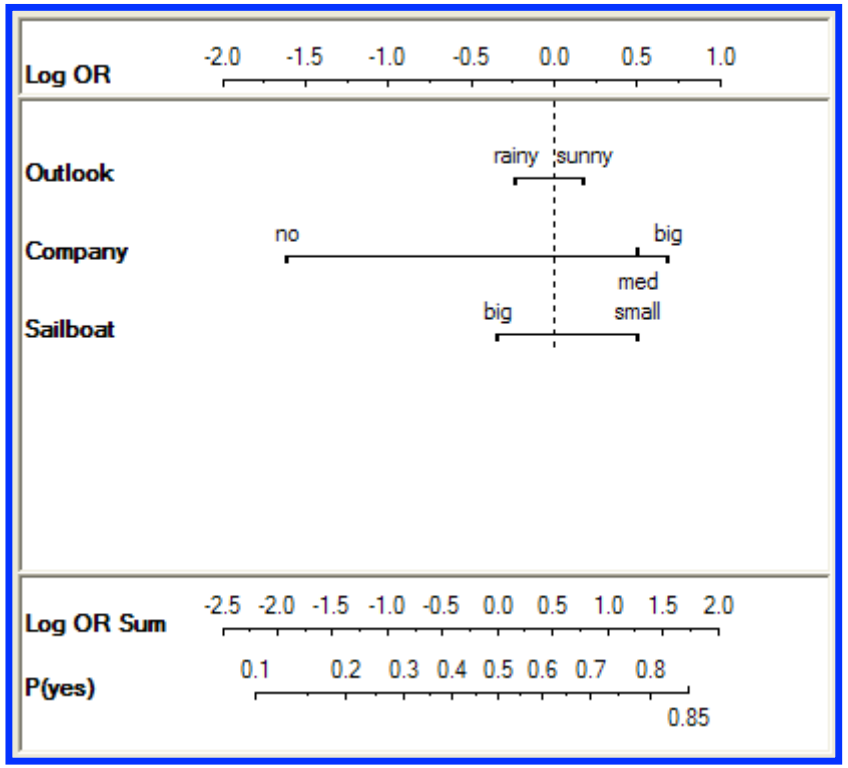
$$p(x_i|e) = \frac{N_{x_i,e}}{N_e}$$

#	Outlook	Company	Sailboat	Sail
1	rainy	big	big	yes
2	rainy	big	small	yes
3	rainy	med	big	yes
4	rainy	med	small	yes
5	sunny	big	big	yes
6	sunny	big	small	yes
7	sunny	med	big	yes
8	sunny	med	big	yes
9	sunny	med	small	yes
10	sunny	no	small	yes
11	sunny	no	big	no
12	rainy	med	big	no
13	rainy	no	big	no
14	rainy	no	big	no
15	rainy	no	small	no
16	rainy	no	small	no
17	sunny	big	big	no
18	sunny	big	small	no
19	sunny	med	big	no
20	sunny	med	big	no



odds	logit
-1.00	0.27
-0.90	0.29
-0.80	0.31
-0.70	0.33
-0.60	0.35
-0.50	0.38
-0.40	0.40
-0.30	0.43
-0.20	0.45
-0.10	0.48
0.00	0.50
0.10	0.53
0.20	0.55
0.30	0.57
0.40	0.60
0.50	0.62
0.60	0.65
0.70	0.67
0.80	0.69
0.90	0.71
1.00	0.73

#	Outlook	Company	Sailboat	Sail
1	rainy	big	big	yes
2	rainy	big	small	yes
3	rainy	med	big	yes
4	rainy	med	small	yes
5	sunny	big	big	yes
6	sunny	big	small	yes
7	sunny	med	big	yes
8	sunny	med	big	yes
9	sunny	med	small	yes
10	sunny	no	small	yes
11	sunny	no	big	no
12	rainy	med	big	no
13	rainy	no	big	no
14	rainy	no	big	no
15	rainy	no	small	no
16	rainy	no	small	no
17	sunny	big	big	no
18	sunny	big	small	no
19	sunny	med	big	no
20	sunny	med	big	no



#	Health	Timing	Complications	Outcome
1	good	bad	some	good
2	good	bad	many	good
3	bad	good	no	good
4	good	good	no	good
5	good	good	no	good
6	bad	good	many	good
7	good	good	many	good
8	bad	good	some	good
9	good	good	some	good
10	good	good	some	good
11	good	good	some	good
12	bad	bad	many	bad
13	good	bad	many	bad
14	good	bad	no	bad
15	good	bad	many	bad
16	good	good	many	bad
17	bad	good	some	bad
18	bad	good	some	bad
19	bad	good	some	bad
20	good	good	some	bad

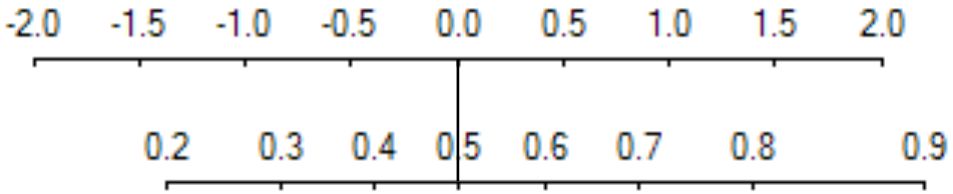
$$f = \log \frac{p(e)}{p(\bar{e})} + \sum_i \log \frac{p(x_i|e)}{p(x_i|\bar{e})}$$

$$p = \frac{1}{1 + \mathrm{e}^{-f}}$$

$$p(e) = \frac{N_e}{N}$$

$$p(x_i|e) = \frac{N_{x_i,e}}{N_e}$$

odds	logit
-1.00	0.27
-0.90	0.29
-0.80	0.31
-0.70	0.33
-0.60	0.35
-0.50	0.38
-0.40	0.40
-0.30	0.43
-0.20	0.45
-0.10	0.48
0.00	0.50
0.10	0.53
0.20	0.55
0.30	0.57
0.40	0.60
0.50	0.62
0.60	0.65
0.70	0.67
0.80	0.69
0.90	0.71
1.00	0.73



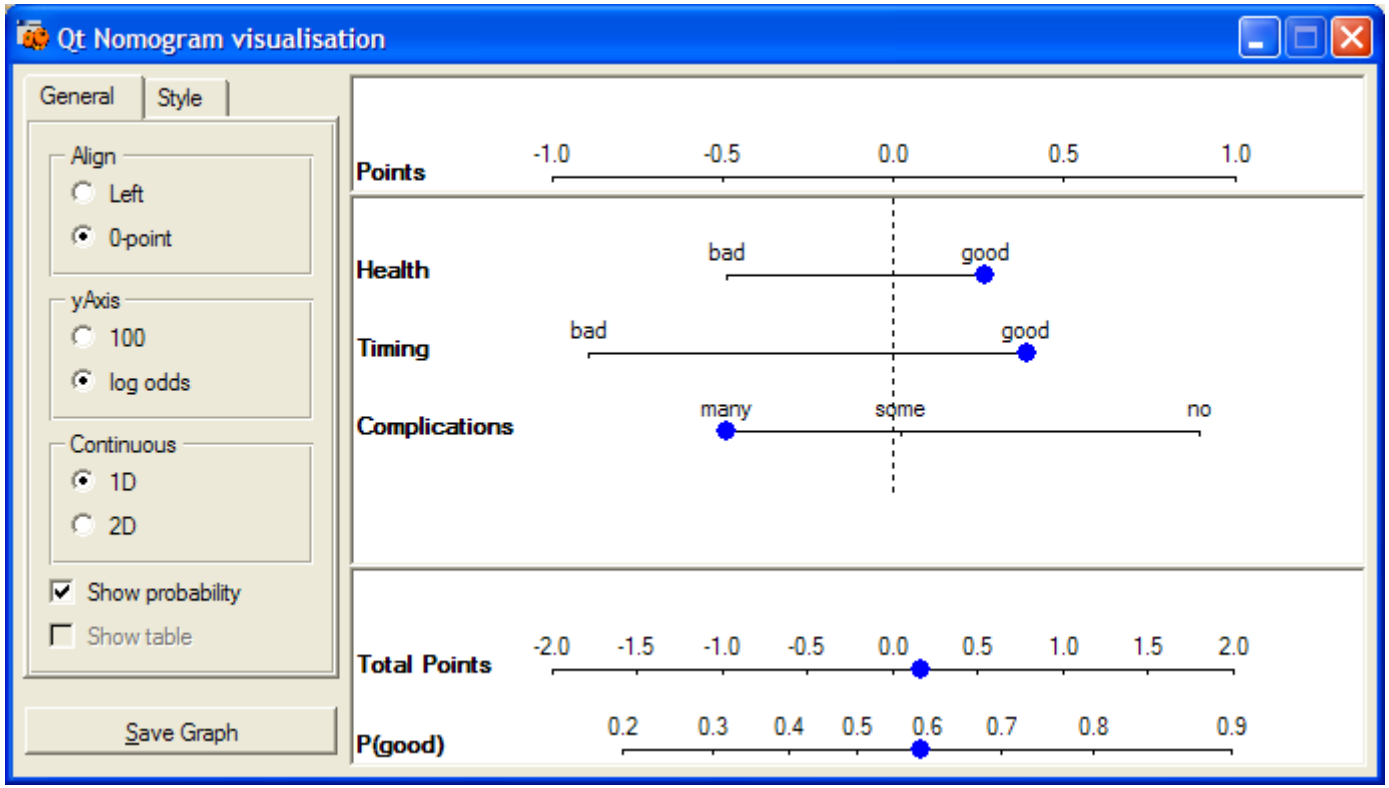
$$p = \frac{1}{1 + \mathrm{e}^{-f}}$$

$$f = \log \frac{p(e)}{p(\bar{e})} + \sum_i \log \frac{p(x_i|e)}{p(x_i|\bar{e})}$$

0.201 logit(0.201)=0.55

log (11/20)/(9/20)=0.201

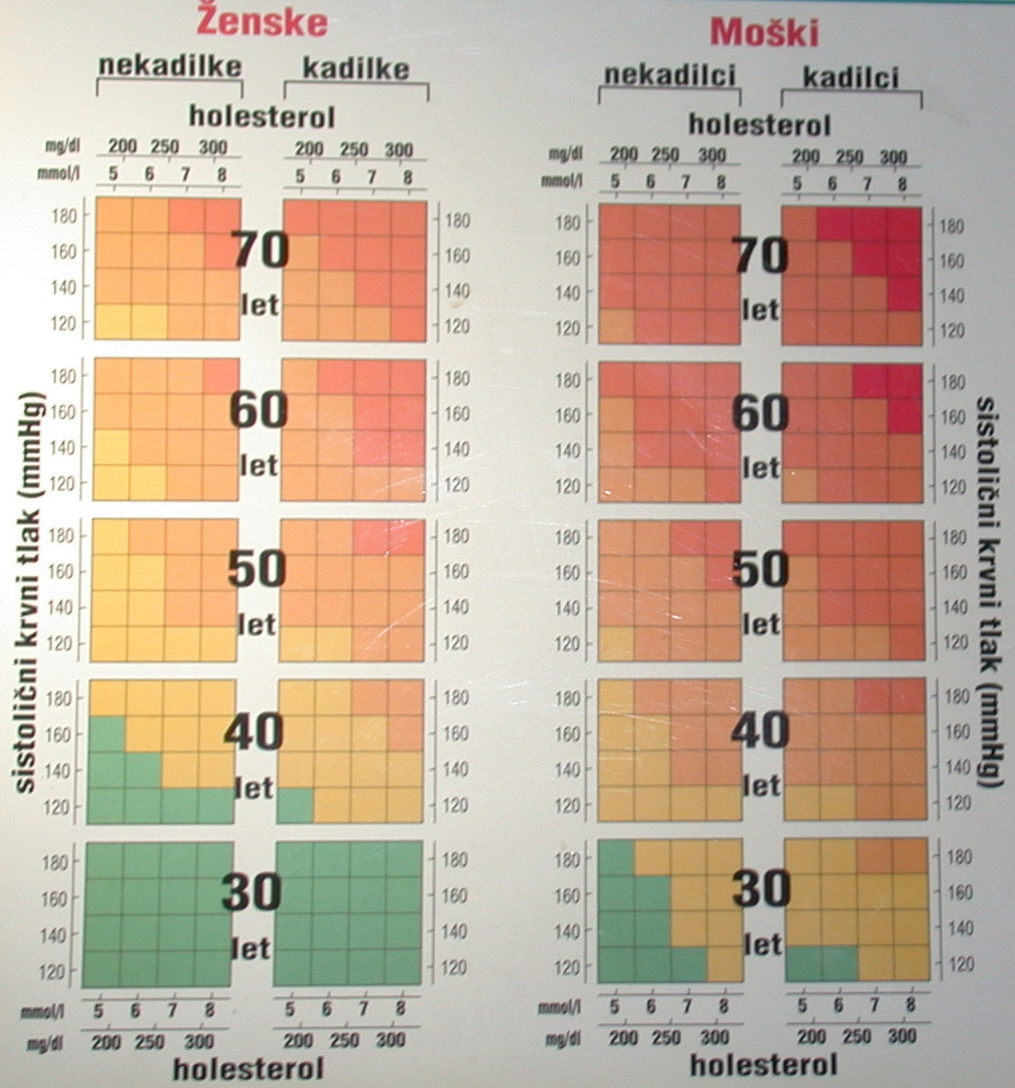
Nomogram


Qt Nomogram visualisation
General
Style
Align
Left
0-point
yAxis
100
log odds
Continuous
1D
2D
Show probability
Show table
Save Graph
Points
-1.0
-0.5
0.0
0.5
1.0
Health
bad
good
Timing
bad
good
Complications
many
some
no
Total Points
-2.0
-1.5
-1.0
-0.5
0.0
0.5
1.0
1.5
2.0
P(good)
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9

Nomogrami

- Nomogrami za klasifikacijo
  - grafična predstavitev verjetnostnega klasifikatorja
  - pomaga pri razumevanju modela (pomembnost atributov)
  - ročno napovedovanje
- Možno je prikazati modele
  - logistične regresije, Cox survival model
  - **naivni Bayesov klasifikator**
- Glavna motivacija: odločanje zdravnikov

TVEGANJE ZA NASTANEK KORONARNE BOLEZNI



KAKO UPORABLJAMO PREGLEDNICO ZA DOLOČANJE TVEGANJA

1. Kako veliko je 10-letno tveganje za pojav koronarnega dogodka (srčna kap) pri neki osebi, določimo tako, da najprej izberemo preglednico, ki ustreza njenemu spolu, kadilskim navadam in starosti.
2. Potem v preglednici poiščemo polje, ki je najbližje vrednostim krvnega tlaka (mmHg) in holesterola pri izbrani osebi.
3. Barvo polja primerjamo z legendo in odčitamo velikost tveganja.
4. Učinek dosmrtne izpostavljenosti dejavnikom tveganja določimo tako, da se znotraj preglednice gibljemo od spodaj navzgor, tj. od manjše k večji starosti.
5. **Opozorilo: pri bolniku s koronarno boleznijo je treba velikost tveganja zvečati za najmanj eno stopnjo.**

   Osebe z družinsko anamnezo koronarnega dogodka v zgodnji dobi, s sladkorno boleznijo ali družinsko anamnezo hiperlipidemije so prav tako bolj ogrožene.

VELIKOST TVEGANJA

Odstotek verjetnosti za pojav nekega koronarnega dogodka v 10 letih

zelo veliko	> 40 %
veliko	20–40 %
zmerno zvečano	10–20 %
blago zvečano	5–10 %
majhno	< 5 %

Temelji na funkciji tveganja, dobljeni v Framinghamski raziskavi, Anderson KM et al. An updated coronary risk profile: A statement for health professionals. *Circulation* 83: 356–362, 1991.

Preglednica temelji na predpostavki, da so vrednosti holesterola HDL 1,0 mmol/l (39 mg/dl) pri moških oziroma 1,1 mmol/l (43 mg/dl) pri ženskah. Osebe z manjšimi vrednostmi in/ali tiste z vrednostmi triglicerídov, večjimi od 2,3 mmol/l (200 mg/dl), so bolj ogrožene.

EVROPSKO KARDIOLOŠKO ZDRUŽENJE

EVROPSKO ZDRUŽENJE PROTI ATEROSKLEROZI

EVROPSKO ZDRUŽENJE PROTI HIPERTENZIJI

Eur Heart J 1994; 15: 1300–31.
Atherosclerosis 1994; 110: 121–61.

Nomogram for Prediction of Final Pathological Stage[1]

Numbers represent the percent probability of the patient having a given final pathological stage based on a logistic regression analysis for all 3 variables combined.

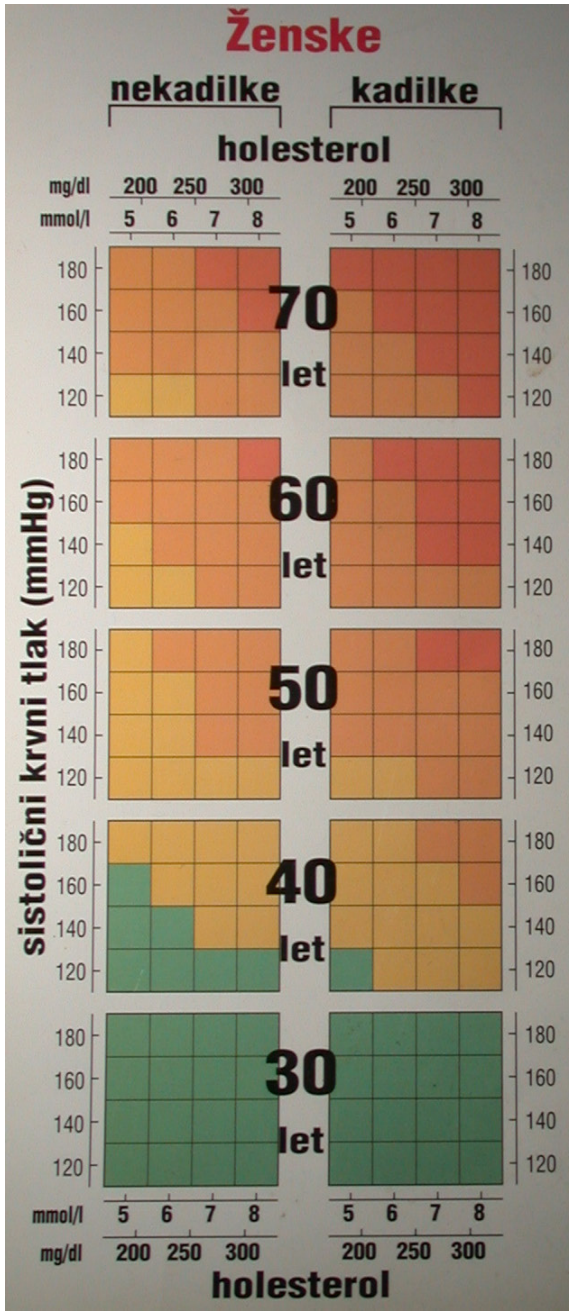
Score	PSA (ng/mL)																												
	0.0–4.0							4.1–10							10.1–20							Greater Than 20						
	Clinical Stage							Clinical Stage							Clinical Stage							Clinical Stage						
	T1a	T1b	T1c	T2a	T2b	T2c	T3a	T1a	T1b	T1c	T2a	T2b	T2c	T3a	T1a	T1b	T1c	T2a	T2b	T2c	T3a	T1a	T1b	T1c	T2a	T2b	T2c	T3a
Prediction of organ-confined disease																													
2–4	100	85	92	88	76	82	—	100	78	82	83	67	71	—	100	—	—	61	52	—	—	—	—	33	20	7	—	—
5	100	78	81	81	67	73	—	100	70	71	73	56	64	43	100	49	55	58	43	37	26	—	—	24	32	—	3	—
6	100	68	69	72	54	60	42	100	53	59	62	44	48	33	—	36	41	44	28	37	19	—	—	22	14	11	4	5
7	—	54	55	61	41	46	—	100	39	43	51	32	37	26	—	24	24	36	19	24	14	—	—	7	18	4	5	3
8–10	—	—	—	48	31	—	—	—	32	31	39	22	25	12	—	11	—	29	14	15	9	—	—	3	3	1	2	2
Prediction of established capsular penetration																													
2–4	0	15	22	14	26	17	—	0	22	29	19	34	27	—	0	—	—	40	49	—	—	—	—	50	80	94	—	—
5	0	22	30	20	34	26	—	0	29	34	28	45	34	58	0	49	40	43	58	61	75	—	—	54	68	—	97	—
6	0	30	34	29	46	38	59	0	45	38	38	56	49	68	—	62	45	56	73	59	82	—	—	53	86	90	96	95
7	—	43	40	39	59	50	—	0	58	44	49	68	59	75	—	73	52	64	81	73	86	—	—	67	80	96	95	98
8–10	—	—	—	50	68	—	—	—	64	48	59	77	71	87	—	87	—	70	86	82	92	—	—	74	97	99	97	98
Prediction of seminal vesicle involvement																													
2–4	0	1	<1	1	2	2	—	0	2	<1	1	3	3	—	0	—	—	3	4	—	—	—	—	<1	12	30	—	—
5	0	3	<1	2	4	4	—	0	4	<1	3	6	6	5	0	7	<1	5	8	12	11	—	—	<1	11	—	29	—
6	0	6	1	5	9	9	8	0	9	1	6	11	12	11	—	15	1	11	19	17	18	—	—	2	35	40	53	31
7	—	12	4	9	17	17	—	0	18	5	12	22	23	18	—	28	6	19	33	33	31	—	—	9	31	73	62	55
8–10	—	—	—	17	29	—	—	—	29	23	22	38	40	40	—	55	—	29	50	53	49	—	—	81	81	93	73	65
Prediction of lymph nodal involvement																													
2–4	0	2	<1	1	2	4	—	0	2	1	1	2	5	—	0	—	—	1	3	—	—	—	—	6	2	7	—	—
5	0	4	1	2	4	8	—	0	4	1	2	5	10	8	0	5	3	2	6	13	11	—	—	9	3	—	29	—
6	0	8	2	3	9	17	15	0	9	2	4	11	19	16	—	11	4	5	13	22	20	—	—	8	9	18	53	31
7	—	15	2	7	18	31	—	0	18	3	8	20	34	28	—	21	7	9	24	39	35	—	—	24	11	44	62	55
8–10	—	—	—	13	32	—	—	—	30	5	15	35	53	50	—	41	—	17	40	59	54	—	—	41	35	76	73	65

Numbers represent probability (%). Dash represents lack of sufficient data to calculate probability.

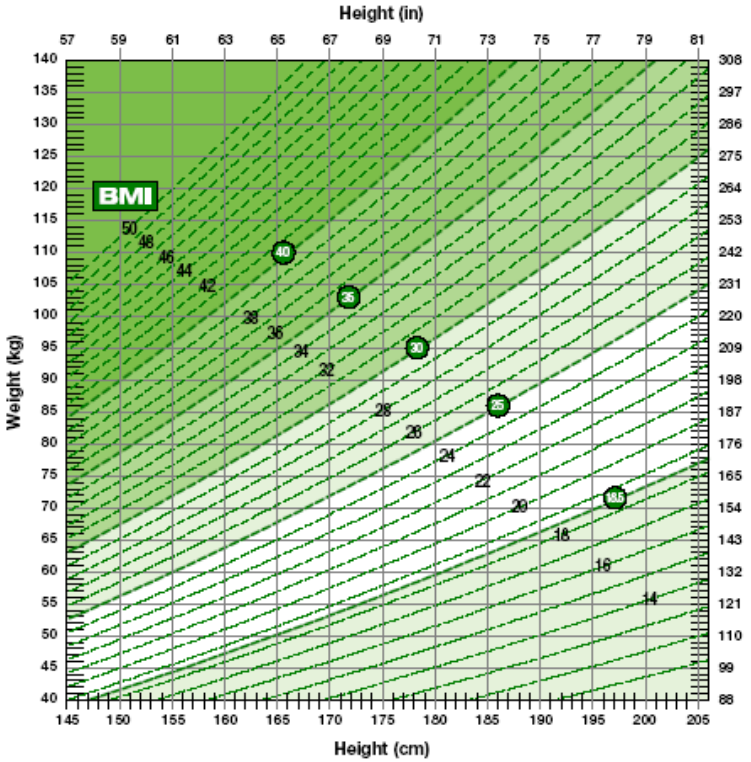
Partin AW et al. J.Urol (1993) 50:110-11.

Tabelarični nomogrami (Lookup Tables)

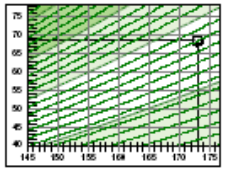
- primerni za diskretne atribute
- majhno število atributov.
- barve pomagajo pri razbiranju trendov



- Body Mass Index (BMI) Nomogram -



For a quick determination of BMI (kg/m^2), use a straight-edge to help locate the point on the chart where height (in or cm) and weight (lb or kg) intersect. **Read the number on the dashed line closest to this point.** For example, an individual who weighs 69 kg and is 173 cm tall has a BMI of approximately 23.



Refer to the table below to identify the level of health risk associated with a particular BMI.

BMI Formula

BMI can also be calculated using this formula
$\text{BMI} = \frac{\text{weight in kilograms}}{(\text{height in metres})^2}$

Note: 1 inch = 2.54 centimetres and 1 pound = 0.45 kilograms

BMI	Risk of developing health problems
< 18.5	Increased
18.5 – 24.9	Least
25.0 – 29.9	Increased
30.0 – 34.9	High
35.0 – 39.9	Very high
≥ 40.0	Extremely high

Note: For persons 65 years and older the 'normal' range may begin slightly above BMI 18.5 and extend into the 'overweight' range.

Adapted from: WHO (2000) Obesity: Preventing and Managing the Global Epidemic: Report of a WHO Consultation on Obesity.

To clarify risk for each individual, other factors such as lifestyle habits, fitness level, and presence or absence of other health risk conditions also need to be considered.

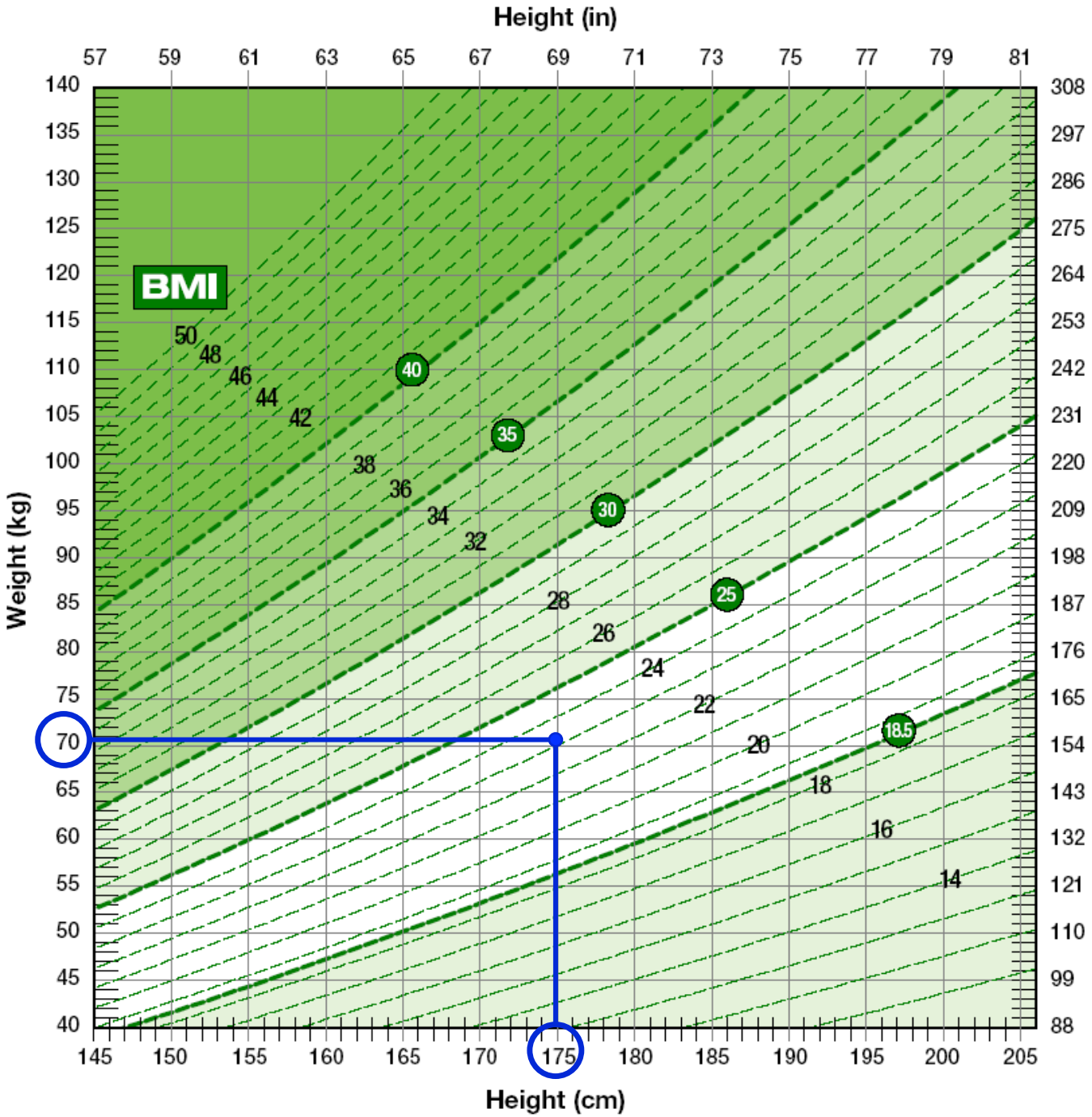
The full report "Canadian Guidelines for Body Weight Classification in Adults", and other resources are available online at:

www.healthcanada.ca/nutrition

ISBN 0-662-33496-5 Cat. No: H49-179/2003-1E Aussi disponible en français

Health Canada Santé Canada

Canada


Height (in)
57 59 61 63 65 67 69 71 73 75 77 79 81
140 135 130 125 120 115 110 105 100 95 90 85 80 75 70 65 60 55 50 45 40
308 297 286 275 264 253 242 231 220 209 198 187 176 165 154 143 132 121 110 99 88
Weight (kg)
BMI
50 48 46 44 42 40 38 36 34 32 30 28 26 24 22 20 18 16 14
40 35 30 25 18.5
145 150 155 160 165 170 175 180 185 190 195 200 205
Height (cm)

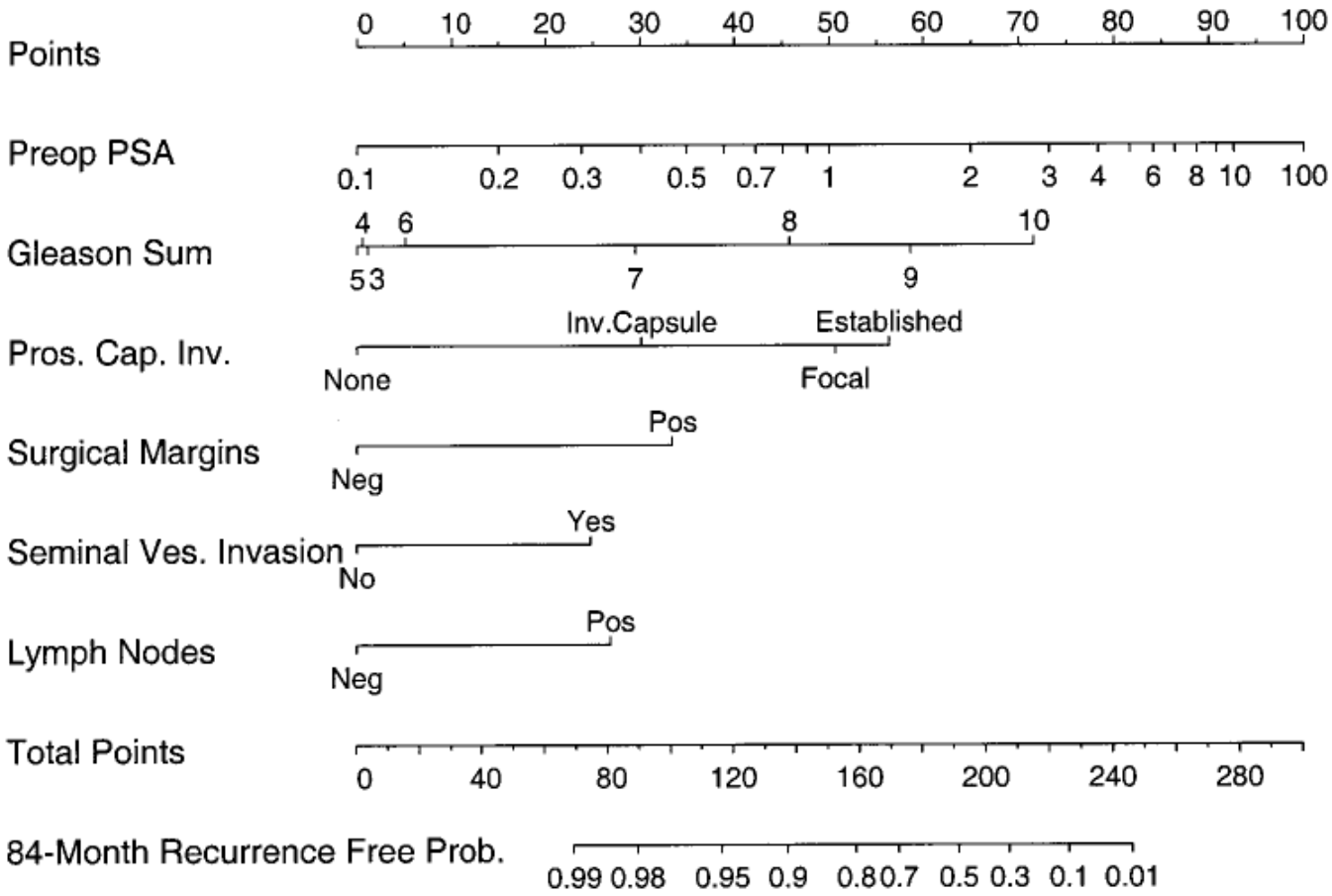
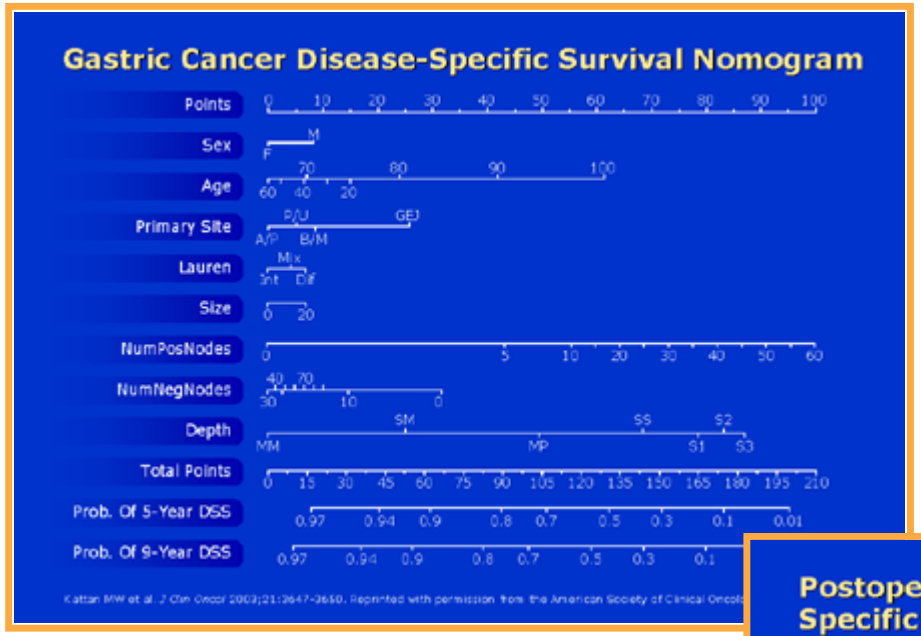


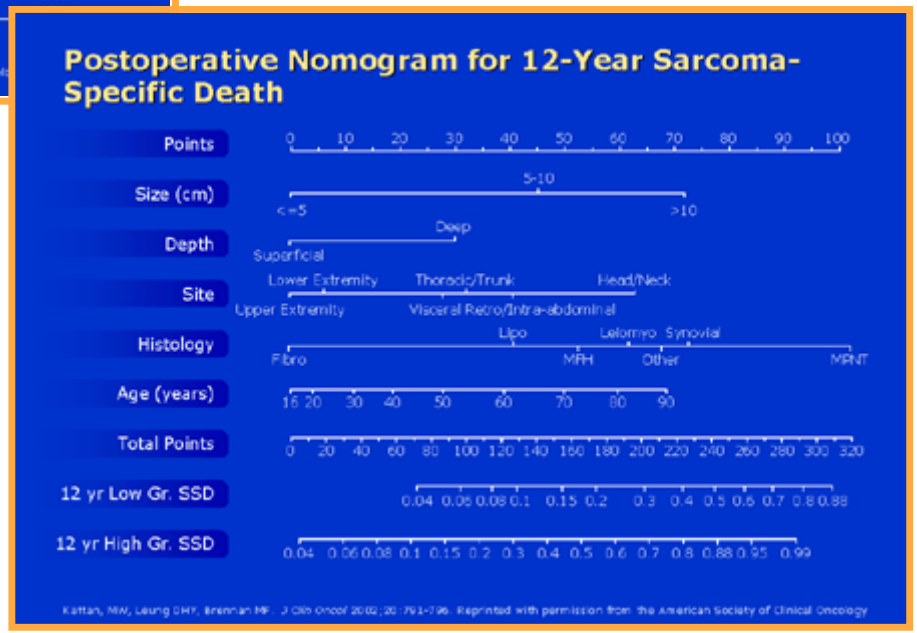
Fig 1. Postoperative nomogram based on 996 patients treated at Baylor College of Medicine, Houston, TX, for predicting PSA recurrence after radical prostatectomy. Data adapted from Kattan et al.[4] Seminal Ves., seminal vesicle; Pros. Cap. Inv., prostatic capsular invasion.

Instructions for Physician: Locate the patient's PSA on the **PSA** axis. Draw a line straight upwards to the **Points** axis to determine how many points towards recurrence the patient receives for his PSA. Repeat this process for the other axes, each time drawing straight upward to the **Points** axis. Sum the points achieved for each predictor and locate this sum on the **Total Points** axis. Draw a line straight down to find the patient's probability of remaining recurrence free for 84 months assuming he does not die of another cause first.

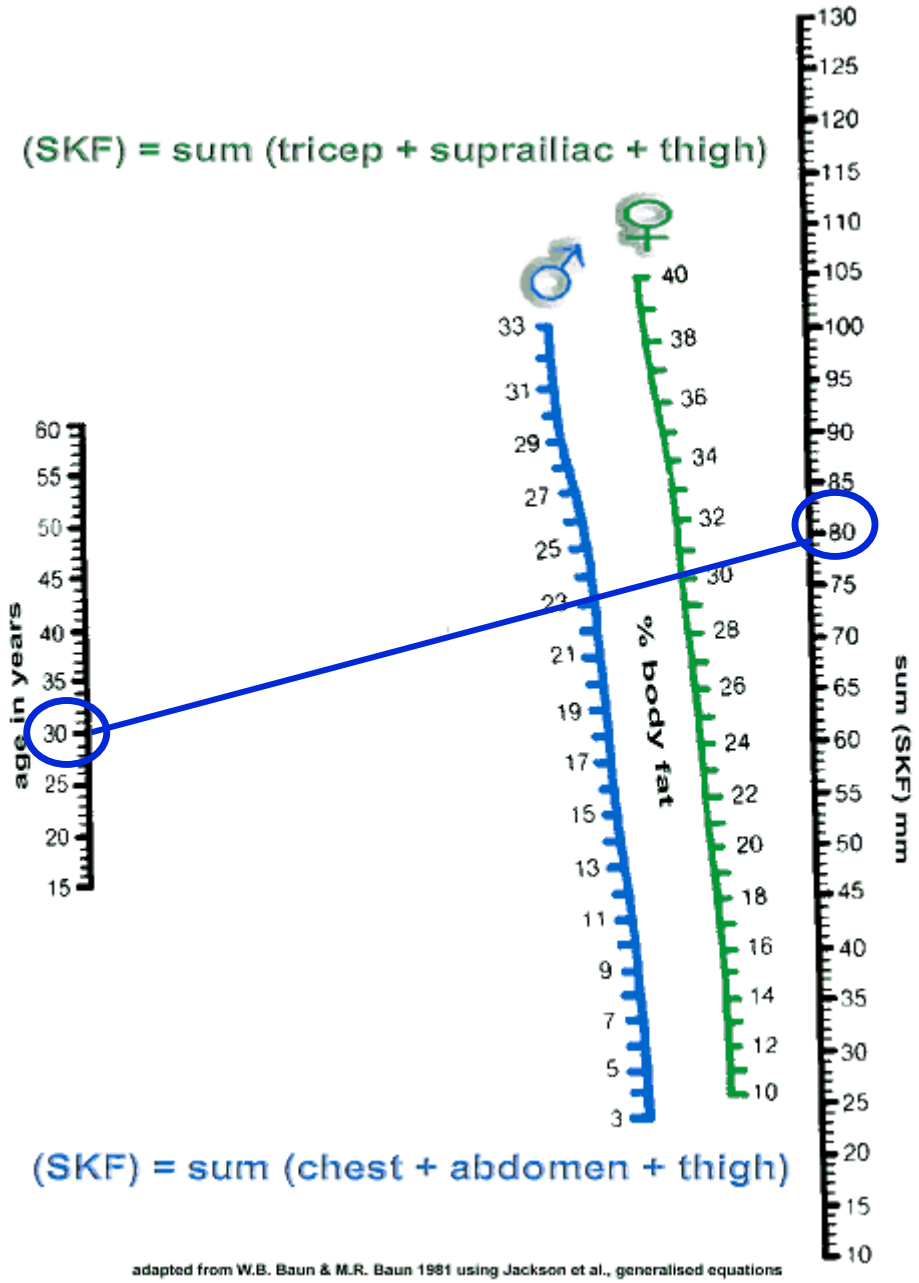
Instruction to Patient: "Mr. X, if we had 100 men exactly like you, we would expect between <predicted percentage from nomogram – 10%> and <predicted percentage + 10%> to remain free of their disease at 7 years following radical prostatectomy, and recurrence after 7 years is very rare."

Grafen, …, Kattan: J Clin Oncol 20:951-956.





http://www.baylorcme.org/nomogram/presentations/kattan/presentation_text.cfm



http://www.insitefitness.com.au/lessons/fitness%20testing/Anthropometry/nomogram.html

Gradnja nomograma

- Podatki z razredom.
- Zgradi napovedni model
  - Naïve Bayesian Classifier
  - Logistic Regression
- Model predstavi z nomogramom
- Uporabi ga
  - analiza modela
  - napoved

#	Outlook	Company	Sailboat	Sail
1	rainy	big	big	yes
2	rainy	big	small	yes
3	rainy	med	big	yes
4	rainy	med	small	yes
5	sunny	big	big	yes
6	sunny	big	small	yes
7	sunny	med	big	yes
8	sunny	med	big	yes
9	sunny	med	small	yes
10	sunny	no	small	yes
11	sunny	no	big	no
12	rainy	med	big	no
13	rainy	no	big	no
14	rainy	no	big	no
15	rainy	no	small	no
16	rainy	no	small	no
17	sunny	big	big	no
18	sunny	big	small	no
19	sunny	med	big	no
20	sunny	med	big	no