

Regressione logistica versus regressione di Cox (Logistic regression versus Cox regression)

Mi chiamo Witold Śmigielski e lavoro in Polonia presso l'Università di Łódź, nell'Istituto di Statistica e Demografia, nonché presso l'Istituto Nazionale di Cardiologia a Varsavia come biostatistico. Oltre all'attività scientifica, lavoro anche come allenatore di squadre giovanili.



Witold Śmigielski - about me

Prima di parlare del tema della mia lezione, vorrei dire qualcosa sul mio paese. A chi pensate quando sentite “Polonia”?

In passato la Polonia era legata soprattutto a:

Papa Giovanni Paolo II,

Lech Wałęsa, leader del movimento anticomunista “Solidarność”,

e per i tifosi italiani anche a Zbigniew Boniek, che ha giocato nella Juventus.

Oggi la Polonia è conosciuta soprattutto per sportivi famosi, come

Robert Lewandowski, Wojciech Szczęsny,

Piotr Zieliński,

e per la nostra campionessa di tennis Iga Świątek.



Polonia

La **regressione logistica** è un metodo statistico usato per analizzare una variabile con **due possibili risultati**, per esempio:

sì / no

malattia / nessuna malattia

successo / insuccesso

vivo / morto

Lo scopo è **calcolare la probabilità che un evento accada.**

Regressione logistica – informazioni di base

Variabile dipendente

Nella regressione logistica la variabile dipendente ha solo due valori:

1 → l'evento accade

0 → l'evento non accade.

Variabili indipendenti

Le variabili che spiegano il fenomeno possono essere:

continue (età, peso),

categoriali (sesso, fumo),

binarie (sì / no).

Variabili

Nella regressione logistica **non modelliamo direttamente i valori 0 e 1**, ma il **logaritmo del rapporto delle probabilità**, chiamato logit:

$$\log \left(\frac{p}{1-p} \right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$$

dove:

- p – è la probabilità che l'evento accada ($Y = 1$),
- β_0 – è il termine costante (intercetta),
- β_i – sono i parametri della regressione, che descrivono l'effetto delle variabili.

Idea del modello (parte matematica)

La regressione lineare:

può dare risultati fuori dall'intervallo da 0 a 1,
non è adatta per variabili con due valori.

La regressione logistica:

dà sempre una probabilità tra 0 e 1,
è facile da interpretare,
è molto usata in medicina e nelle scienze sociali.

Perché non usare la regressione lineare?

I risultati sono spesso interpretati con l'**odds ratio (OR)**.

$OR > 1 \rightarrow$ la probabilità dell'evento aumenta

$OR < 1 \rightarrow$ la probabilità diminuisce

$OR = 1 \rightarrow$ nessun effetto

Esempio: $OR = 1,3$ significa che la probabilità aumenta del 30%.

Interpretazione dei risultati

La regressione logistica è **usata molto spesso in:**
medicina e biostatistica,
epidemiologia,
scienze sociali,
economia,
analisi dei dati.

► Dove si usa la regressione logistica?

Esempio



La nostra analisi

Il più grande studio epidemiologico in Polonia si chiama WOBASZ. La prima edizione è stata fatta negli anni 2003–2005. Dopo 20 anni abbiamo controllato se i partecipanti sono vivi o morti.

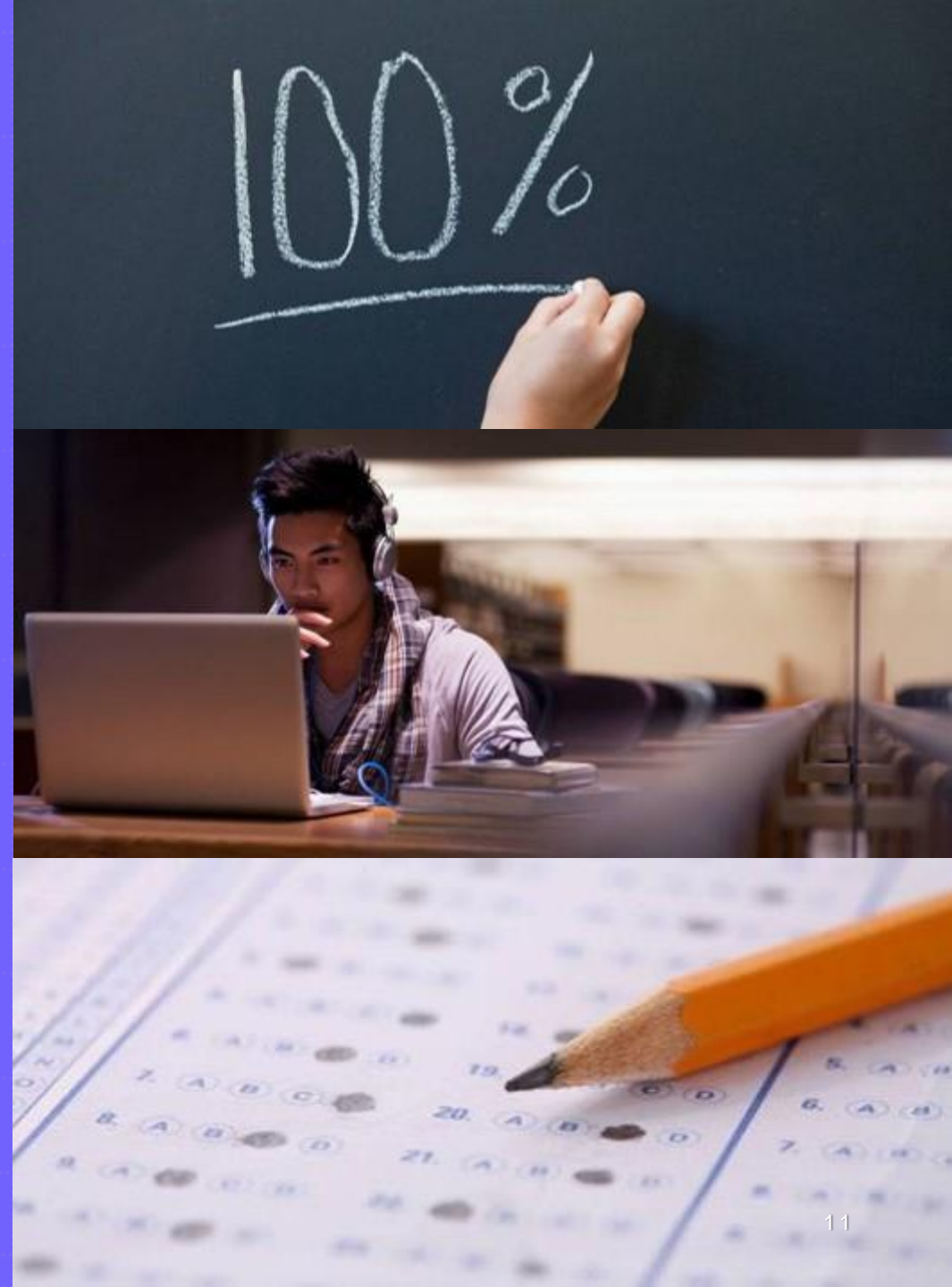
Numero totale di persone: 14 067

Metodo: regressione logistica

Variabile dipendente: stato di sopravvivenza

1 – persona morta

0 – persona viva



Come stare in buona salute?

Prima di iniziare l'analisi, parliamo insieme di come stare in buona salute? **Per favore, rispondete in inglese.**

- a) Mangiare frutta e verdura
- b) Essere fisicamente attivi
- c) Non fumare
- d) Non abusare di alcol
- e) Consumare meno zucchero
- f) Dormire bene
- g) Ridurre lo stress
- h) Avere buoni rapporti con la famiglia e gli amici
- i) Altri fattori ...

**Ma... abbiamo prove che questo
fa davvero bene alla nostra
salute?**

In Polonia, l'Istituto Nazionale di Cardiologia propone "10 per il cuore".

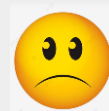
- 1) **Pressione sanguigna ottimale**
- 2) **Livello ottimale di colesterolo**
- 3) **Livello ottimale di glucosio**
- 4) **Peso corporeo ottimale**
- 5) **Dieta adeguata**
- 6) **Non fumare**
- 7) **Non abusare di alcol**
- 8) **Sonno sano**
- 9) **Basso livello di stress**
- 10) **Attività fisica**



Come lo misuriamo nella pratica?

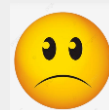
Pressione sanguigna ottimale (sistolica/diastolica)

< 120/70 [0 punti] / 120-139 / 70-80 [1 punto] / >= 140/90 [2 punti]



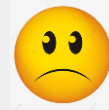
Livello ottimale di colesterolo

< 115 [0 punti] / 115–189 [1 punto] / >= 190 [2 punti]



Livello ottimale di glucosio [mg/dL]

< 100 [0 punti] / 100-125 [1 punto] / >= 126 [2 punti]



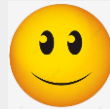
Peso corporeo ottimale

In base al BMI:

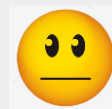
$$BMI = \frac{\text{Peso [kg]}}{(\text{altezza [m]})^2}$$

Valore del BMI:

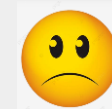
<25 [0 punti]



25-30 [1 punto]



>30 [2 punti]



In base alla circonferenza della vita:

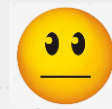
<80 (donne) / < 94 (uomini)

[0 punti]



80-87 (donne) / 94-101 (uomini)

[1 punto]



> 87 (donne) / > 101 (uomini)

[2 punti]



Possiamo sempre fidarci del BMI?

Secondo il BMI, quest'uomo è obeso...



In alcuni casi dobbiamo basarci sulle dichiarazioni delle persone intervistate:

Consumo di alcol:

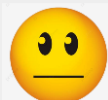
Per niente

[0 punti]



=< 100 g (ethanol) / settimana

[1 punto]



>100 (ethanol) / settimana

[2 punti]



Fumare:

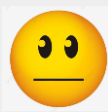
Mai fumatore

[0 punti]



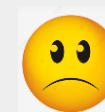
Ex fumatore

[1 punto]



Fumatore attuale

[2 punti]



Attività fisica nel tempo libero:

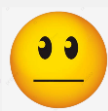
2–3 volte alla settimana

[0 punti]



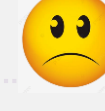
Una volta alla settimana

[1 punto]



Per niente

[2 punti]

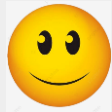


Alcune variabili sono difficili da misurare

Stress:

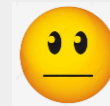
Non più di prima

[0 punti]



Più di prima

[1 punto]



Sempre stressato

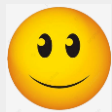
[2 punti]



Qualità del sonno

Alta

[0 punti]



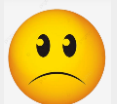
Non male, ma peggiore di prima

[1 punto]



Disturbo grave del sonno

[2 punti]



Usando la regressione logistica vogliamo verificare se i singoli elementi del “10 per il cuore” influenzano in modo statisticamente significativo la durata della vita dei partecipanti allo studio. Il periodo di osservazione è di 20 anni (follow-up di 20 anni).

Variabile dipendente:

0 - ha vissuto almeno 20 anni

1 - è morto durante il periodo di follow-up di 20 anni.

Variabili indipendenti:

gli elementi del “10 per il cuore”.

La nostra analisi

Fattori demografici

Oltre alle variabili già menzionate, la sopravvivenza delle persone studiate dipende in modo evidente da:

- sesso
- età





Uomini

50–64 anni

Analisi univariata

Pressione sanguigna ottimale

N=1864	Constant	Blood press.
Value	-0,5826383	0,2566431
Bład standard.	0,1126125	0,06582948
t(1862)	-5,173834	3,898605
p	0,0000002540494	0,0001001775
-95%CL	-0,8034983	0,1275358
+95%CL	-0,3617783	0,3857504
Chi-kwadrat Walda	26,76856	15,19912
p	0,0000002307204	0,0000969459
Odds ratio	0,5584232	1,292584
-95%CL	0,4477598	1,136026
+95%CL	0,6964368	1,470718

Livello ottimale di colesterolo

N=1763	Constant	Cholesterol
Value	-0,1820787	-0,02867843
Bład standard.	0,0839121	0,08286213
t(1761)	-2,169874	-0,3460982
p	0,03014967	0,7293103
-95%CL	-0,3466566	-0,1911969
+95%CL	-0,01750091	0,1338401
Chi-kwadrat Walda	4,708355	0,1197839
p	0,03002368	0,7292712
Odds ratio	0,8335357	0,9717289
-95%CL	0,7070481	0,8259699
+95%CL	0,9826514	1,14321

OR = 1,29

For participants **with higher blood pressure (120-139 / 70-80)** in compare to participants with **optimal blood pressure (<120 / <70)**, the risk of death during 20-years period **is 29% higher**.

Also for participants with **to high blood pressure (140+ / 80+)** in compare to participants with **higher blood pressure (120-139 / 70-80)** , the risk of death during 20-years period **is 29% higher**.

But for participants with **to high blood pressure (140+ / 80+)** in compare to participants with **optimal blood pressure (<120 / <70)**, the risk of death during 20-years period **is 66% higher!**

Why? 0 vs. 1 (diference in variable's value is 1), 1 vs. 2 (diference in variable's value is 1), but 0 vs. 2 (diference in variable's value is 2), so:

$$1,29 * 1,29 = 1,6641$$

Pressione sanguigna ottimale

Analisi univariata

Livello ottimale di glucosio

N=1823	Constant	Glucose
Value	-0,2965835	0,234219
Błąd standard.	0,05464643	0,06451843
t(1821)	-5,427318	3,630265
p	0,00000006487385	0,0002909523
-95%CL	-0,4037598	0,1076811
+95%CL	-0,1894072	0,3607569
Chi-kwadrat Walda	29,45577	13,17882
p	0,00000005762234	0,0002835718
Odds ratio	0,7433535	1,263921
-95%CL	0,6678045	1,113693
+95%CL	0,8274495	1,434415

Peso corporeo ottimale

N=1843	Constant	Overweight
Value	-0,10222	-0,08799013
Błąd standard.	0,07824542	0,06269404
t(1841)	-1,306403	-1,403485
p	0,1915789	0,1606409
-95%CL	-0,2556791	-0,210949
+95%CL	0,05123908	0,03496876
Chi-kwadrat Walda	1,706688	1,96977
p	0,1914251	0,160482
Odds ratio	0,9028309	0,9157699
-95%CL	0,7743904	0,8098153
+95%CL	1,052575	1,035587

Analisi univariata

Fumare

N=1834	Constant	Smoking
Value	-0,8873461	0,6129604
Bład standard.	0,08862941	0,06253055
t(1832)	-10,01187	9,802575
p	5,21201300E-23	3,80419500E-22
-95%CL	-1,061171	0,4903217
+95%CL	-0,7135208	0,735599
Chi-kwadrat Walda	100,2375	96,09047
p	1,46414900E-23	1,18120500E-22
Odds ratio	0,411747	1,845888
-95%CL	0,3460502	1,632841
+95%CL	0,4899163	2,086732

Non abusare di alcol

N=1861	Constant	Alcohol
Value	-0,3033679	0,1056094
Bład standard.	0,110651	0,08726397
t(1859)	-2,741664	1,210229
p	0,006171449	0,2263449
-95%CL	-0,5203812	-0,0655363
+95%CL	-0,08635455	0,2767551
Chi-kwadrat Walda	7,51672	1,464654
p	0,006116282	0,2262
Odds ratio	0,7383274	1,111388
-95%CL	0,594294	0,936565
+95%CL	0,9172689	1,318843

Analisi univariata

Sonno sano

N=1686	Constant	Sleeping
Value	-0,2913307	0,1517228
Błąd standard.	0,06670533	0,07291596
t(1684)	-4,367427	2,08079
p	0,00001333774	0,03760394
-95%CL	-0,4221648	0,00870735
+95%CL	-0,1604966	0,2947383
Chi-kwadrat Walda	19,07442	4,329688
p	0,00001261132	0,03746105
Odds ratio	0,7472685	1,163838
-95%CL	0,655626	1,008745
+95%CL	0,8517208	1,342775

Stress

N=1685	Constant	Stress
Value	-0,3006726	0,1972505
Błąd standard.	0,06534699	0,08136533
t(1683)	-4,601171	2,424258
p	0,000004515634	0,01544445
-95%CL	-0,4288425	0,03766265
+95%CL	-0,1725027	0,3568384
Chi-kwadrat Walda	21,17077	5,877027
p	0,000004217333	0,01534521
Odds ratio	0,7403201	1,218049
-95%CL	0,6512625	1,038381
+95%CL	0,841556	1,428805

Analisi univariata

Attività fisica

N=1847	Constant	Phys. Activity
Value	-0,3085157	0,1505342
Błąd standard.	0,06338739	0,05113183
t(1845)	-4,867146	2,94404
p	0,000001228569	0,003280076
-95%CL	-0,4328342	0,05025183
+95%CL	-0,1841971	0,2508165
Chi-kwadrat Walda	23,68911	8,667374
p	0,000001137603	0,003241905
Odds ratio	0,7345365	1,162455
-95%CL	0,648668	1,051536
+95%CL	0,8317719	1,285074

Analisi univariate uomini, 50-64 anni

- 1) **Pressione sanguigna ottimale**
- 2) **Livello ottimale di colesterolo**
- 3) **Livello ottimale di glucosio**
- 4) **Peso corporeo ottimale**
- 5) **Dieta adeguata**
- 6) **Non fumare**
- 7) **Non abusare di alcol**
- 8) **Sonno sano**
- 9) **Basso livello di stress**
- 10) **Attività fisica**



Analisi multivariata (fase 1)

N=1601	Constant	Blood pressure	Glucose	Smoking	Phys. Activity	Sleeping	Stress
Value	-1,575157	0,2524785	0,214506	0,6318752	0,07622429	0,03370324	0,1468818
Błąd standard.	0,1711526	0,07576438	0,07192609	0,06841654	0,0572801	0,08179701	0,09140583
t(1594)	-9,203228	3,332417	2,982311	9,235709	1,330729	0,4120351	1,606919
p	1,05670300E-19	0,0008806622	0,002904233	7,92618100E-20	0,1834685	0,6803691	0,1082701
-95%CL	-1,910864	0,1038702	0,07342629	0,4976794	-0,03612797	-0,1267378	-0,0324065
+95%CL	-1,239449	0,4010868	0,3555857	0,7660711	0,1885765	0,1941443	0,3261701
Chi-kwadrat Walda	84,69941	11,105	8,894179	85,29833	1,77084	0,1697729	2,582189
p	3,67825100E-20	0,000861928	0,002862966	2,71919000E-20	0,1832877	0,6803164	0,108082
Odds ratio	0,2069751	1,287212	1,23925	1,881135	1,079205	1,034278	1,158217
-95%CL	0,1479524	1,109456	1,076189	1,6449	0,9645169	0,8809646	0,9681129
+95%CL	0,2895438	1,493447	1,427016	2,151297	1,20753	1,214271	1,385651

Analisi multivariata (fase 2)

N=1602	Constant	Blood pressure	Glucose	Smoking	Phys. Activity	Stress
Value	-1,569126	0,2557811	0,2173437	0,6338933	0,07788973	0,1573889
Bład standard.	0,1697979	0,0755358	0,07176095	0,06839326	0,05724163	0,08664213
t(1596)	-9,24114	3,386224	3,028719	9,268359	1,360718	1,81654
p	7,54312700E-20	0,0007258184	0,002495334	5,92296300E-20	0,1737949	0,06947511
-95%CL	-1,902177	0,1076213	0,07658809	0,4997433	-0,03438696	-0,01255547
+95%CL	-1,236076	0,4039409	0,3580993	0,7680434	0,1901664	0,3273332
Chi-kwadrat Walda	85,39867	11,46651	9,173137	85,90248	1,851554	3,299817
p	2,58499400E-20	0,0007094676	0,00245789	2,00496000E-20	0,1736123	0,06929695
Odds ratio	0,2082271	1,29147	1,242771	1,884935	1,081003	1,170451
-95%CL	0,1492434	1,113626	1,079597	1,648298	0,9661976	0,987523
+95%CL	0,290522	1,497715	1,430608	2,155545	1,209451	1,387264

Analisi multivariata (fase 3)

N=1618	Constant	Blood pressure	Glucose	Smoking	Stress
Value	-1,534033	0,2622478	0,2215644	0,6462889	0,1554299
Bład standard.	0,1649073	0,07511631	0,07101574	0,06800843	0,08616245
t(1613)	-9,302395	3,491223	3,119933	9,503071	1,803916
p	4,32240100E-20	0,000493767	0,001840951	7,08356400E-21	0,07143091
-95%CL	-1,857488	0,114912	0,08227155	0,5128947	-0,01357227
+95%CL	-1,210578	0,4095837	0,3608572	0,7796831	0,324432
Chi-kwadrat Walda	86,53455	12,18864	9,733985	90,30835	3,254114
p	1,45770800E-20	0,0004814618	0,001810526	2,17476700E-21	0,07125381
Odds ratio	0,2156642	1,299849	1,248028	1,908445	1,16816
-95%CL	0,1560642	1,121775	1,085751	1,670119	0,9865194
+95%CL	0,2980251	1,506191	1,434559	2,180781	1,383245

Analisi multivariata (finale)

N=1793	Constant	Blood pressure	Glucose	Smoking
Value	-1,474587	0,2821141	0,2198203	0,6498115
Bład standard.	0,149812	0,07040747	0,06791634	0,06467949
t(1789)	-9,842917	4,006877	3,236633	10,04664
p	2,67820700E-22	0,00006404948	0,001231603	3,85312700E-23
-95%CL	-1,768412	0,1440245	0,08661655	0,5229562
+95%CL	-1,180762	0,4202036	0,353024	0,7766668
Chi-kwadrat Walda	96,88303	16,05506	10,47579	100,935
p	7,92498000E-23	0,00006166602	0,001210719	1,03072700E-23
Odds ratio	0,2288732	1,32593	1,245853	1,91518
-95%CL	0,1706037	1,154912	1,090478	1,687007
+95%CL	0,3070446	1,522271	1,423365	2,174213

Seconda
soluzione.

Uso dell'età
come variabile
standardizzata.



Analisi (multi)univariata

Pressione sanguigna ottimale

N=6609	Constant	Age	Blood pressure
Value	-6,540457	0,1097796	0,1900011
Błąd standard.	0,163056	0,00302822	0,04682564
t(6606)	-40,11171	36,25217	4,057629
p	0	0	0,00005014491
-95%CL	-6,8601	0,1038433	0,09820767
+95%CL	-6,220815	0,1157158	0,2817944
Chi-kwadrat Walda	1608,949	1314,22	16,46435
p	0	0	0,00004969105
Odds ratio	0,001443828	1,116032	1,209251
-95%CL	0,001048809	1,109426	1,103192
+95%CL	0,001987625	1,122677	1,325506

Univariate analysis – males (aged standardised)

- 1) **Optimal blood pressure**
- 2) **Optimal cholesterol levels**
- 3) **Optimal blood glucose levels**
- 4) **Optimal body weight**
- 5) **No smoking**
- 6) **No alcohol consumption**
- 7) **Healthy sleep**
- 8) **Low stress levels**
- 9) **Physical activity**



Analisi multivariata (fase 1 e finale)

[uomini; età come variabile standardizzata]

N=5577	Constant	Age	Blood pressure	Glucose	Smoking	Overweight	Phys. Activity	Sleeping	Stress
Value	-7,720126	0,1153801	0,1845316	0,2679292	0,1157031	0,5111197	0,1527866	0,1231267	0,1940558
Bład standard.	0,2232886	0,003613647	0,05454281	0,0571199	0,05272721	0,04888865	0,04039763	0,05919633	0,06335858
t(5568)	-34,57465	31,92899	3,383243	4,690645	-2,194372	10,45477	3,78207	2,079973	3,062818
p	0	0	0,0007212935	0,000002788534	0,02824984	2,38901800E-25	0,0001571597	0,03757367	0,002203105
-95%CL	-8,157859	0,1082959	0,07760642	0,1559519	-0,219069	0,4152789	0,07359154	0,007078859	0,06984829
+95%CL	-7,282393	0,1224643	0,2914568	0,3799065	-0,01233722	0,6069605	0,2319817	0,2391746	0,3182634
Chi-kwadrat Walda	1195,406	1019,461	11,44634	22,00215	4,815269	109,3022	14,30405	4,326286	9,380856
p	0	0	0,0007172096	0,000002734686	0,02821583	1,53193400E-25	0,0001558125	0,03753597	0,002194452
Odds ratio	0,0004438046	1,1223	1,202655	1,307255	0,8907396	1,667157	1,165076	1,131028	1,214164
-95%CL	0,0002864751	1,114377	1,080697	1,16877	0,8032663	1,514793	1,076367	1,007104	1,072345
+95%CL	0,000687538	1,130279	1,338376	1,462148	0,9877385	1,834846	1,261097	1,2702	1,374738

Univariate analysis – comparison of variable's significance [red, $p < 0,05$]

Variable	Male [aged 50 - 64]	Male [aged standarised]	Female [aged 50 - 64]	Female [aged standarised]
Blood pressure				
Cholesterol				
Glucose				
Overweight				
Smoking				
Alcohol				
Phys. Activity				
Sleeping				
Stress				

Donne



Analisi multivariata (fase 1 e finale)

[donne; età come variabile standardizzata]

N=6027	Stala B0	Age	Blood pressure	Glucose	Smoking	Alcohol	Phys. Activity	Stess
Value	-8,4287	0,1204879	0,1426866	0,391035	0,4183753	-0,3877879	0,1798375	0,1409899
Bład standard.	0,2952905	0,004688716	0,05470673	0,06356198	0,0536697	0,08247155	0,04391161	0,06399286
t(6019)	-28,54376	25,69743	2,608209	6,152026	7,795373	-4,702081	4,095442	2,203213
p	0	0	0,009124223	0,000000008142058	7,51605400E-15	0,000002632671	0,0000426888	0,02761745
-95%CL	-9,007575	0,1112964	0,03544181	0,2664307	0,3131635	-0,5494616	0,09375499	0,01554101
+95%CL	-7,849825	0,1296795	0,2499314	0,5156392	0,5235872	-0,2261141	0,26592	0,2664388
Chi-kwadrat Walda	814,7463	660,3579	6,802754	37,84743	60,76785	22,10956	16,77265	4,854149
p	0	0	0,009105963	0,0000000007740465	6,61577200E-15	0,000002585965	0,00004223983	0,02758681
Odds ratio	0,0002185053	1,128047	1,153368	1,47851	1,519491	0,6785563	1,197023	1,151413
-95%CL	0,0001224785	1,117726	1,036077	1,305297	1,367745	0,5772605	1,098291	1,015662
+95%CL	0,0003898201	1,138463	1,283937	1,674709	1,688072	0,7976271	1,304631	1,305308

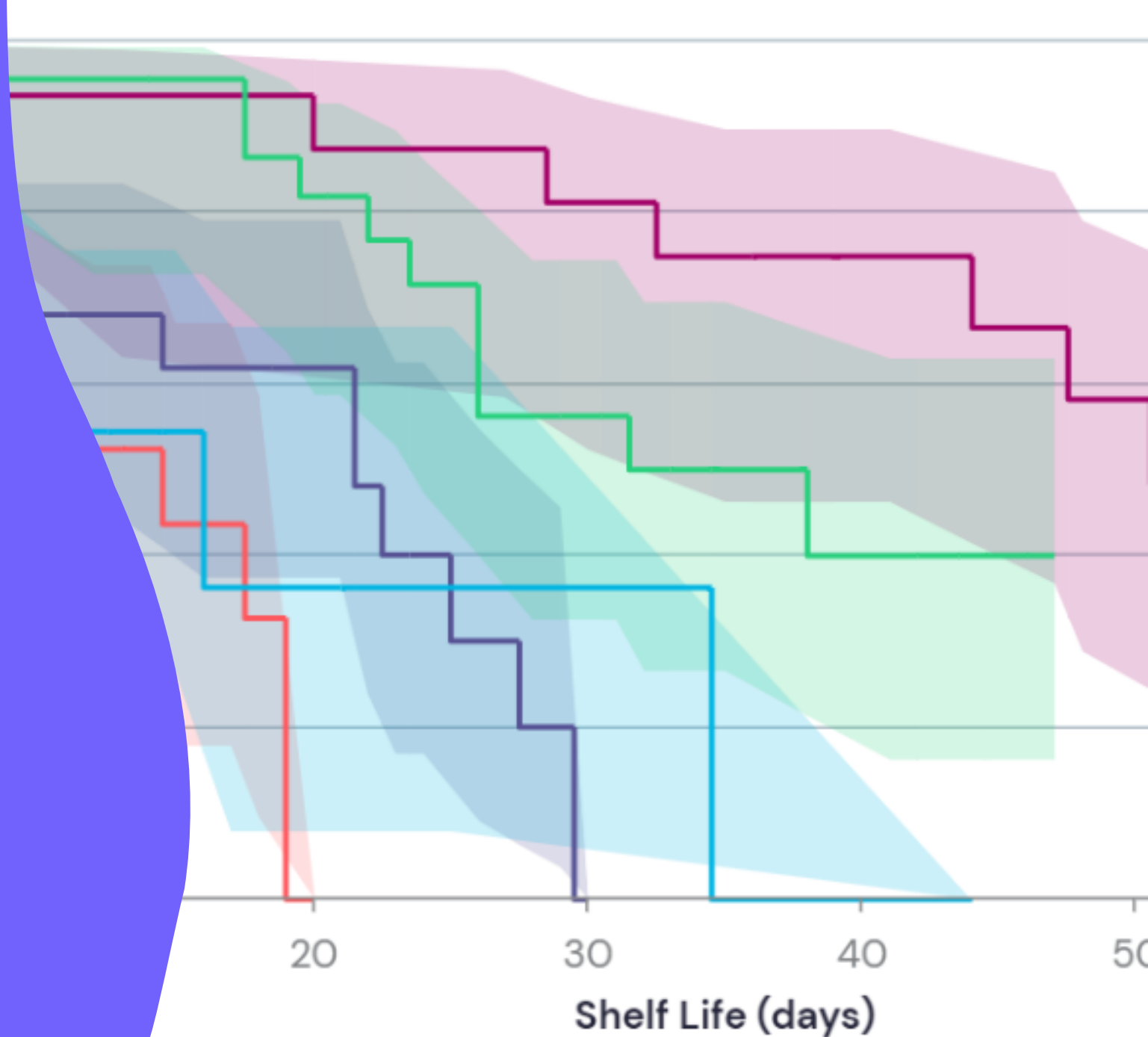
Punti deboli della regressione logistica nell'analisi di sopravvivenza (20 anni di follow-up)

Se una persona:

- a) **muore all'inizio del periodo** analizzato (dopo 1 settimana) – segniamo solo che è morta (codice 1)
- b) **muore a metà del periodo analizzato** (dopo 10 anni) – segniamo solo che è morta (codice 1)
- c) **muore alla fine del periodo analizzato** (dopo 19 anni) – segniamo solo che è morta (codice 1)

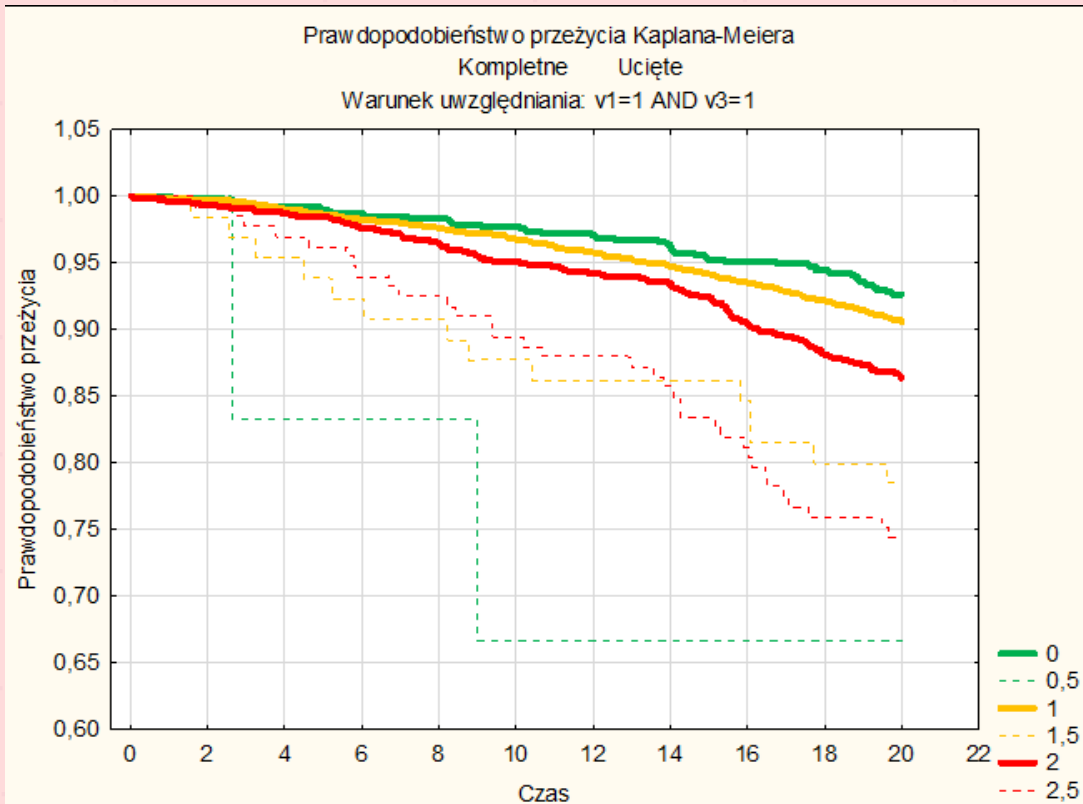
Questa analisi **distingue solo tra vivo e morto**. Non analizza **quanto a lungo una persona ha vissuto**.

Analizziamo
i dati di
sopravvivenza
delle variabili
analizzate –
curve di Kaplan-
Meier

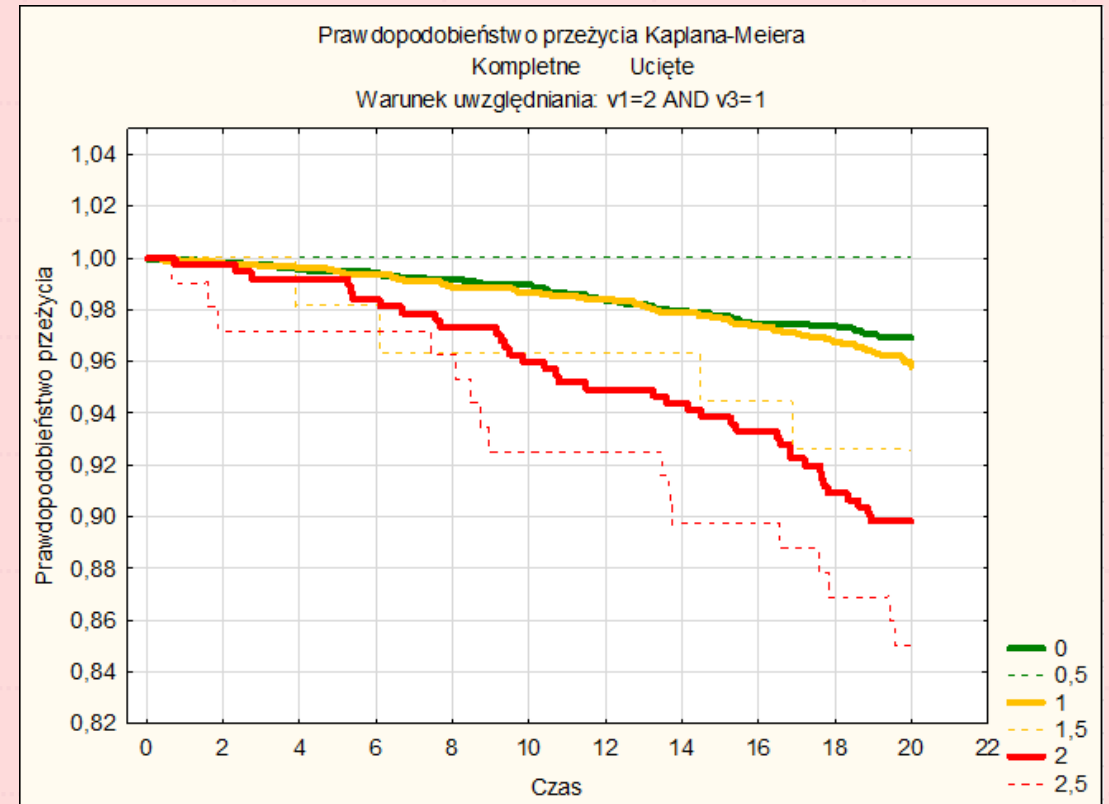


Blood pressure (<50 years old)

Male

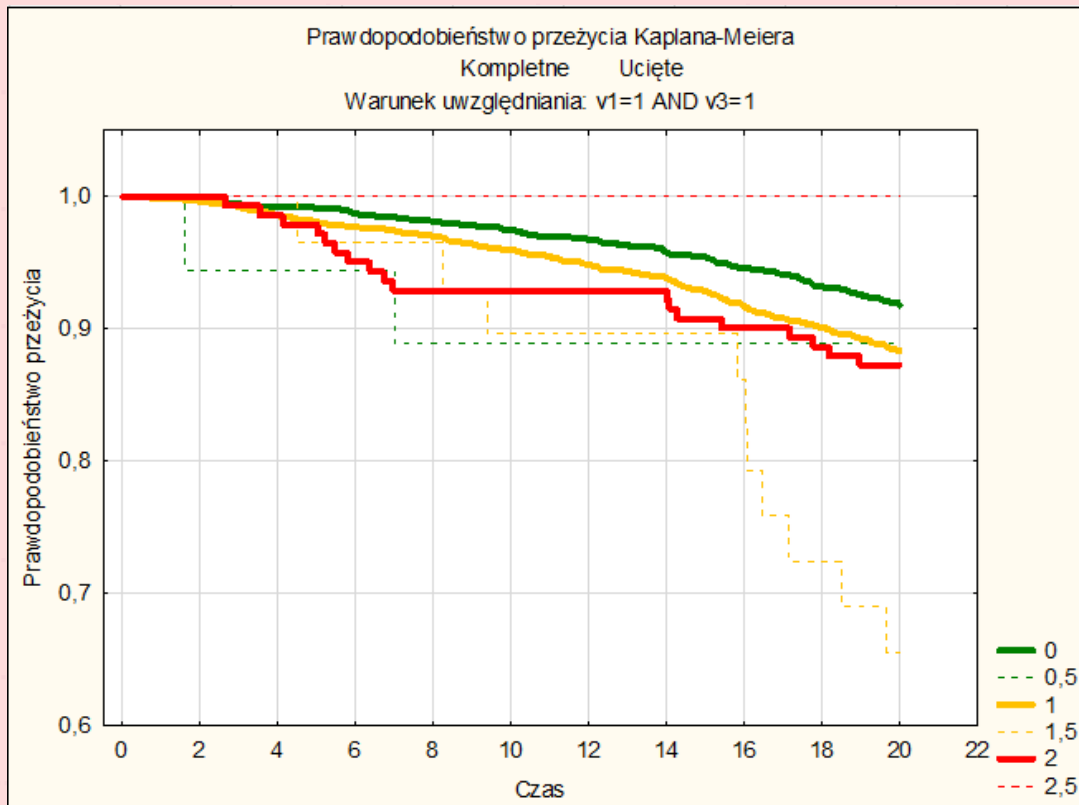


Female

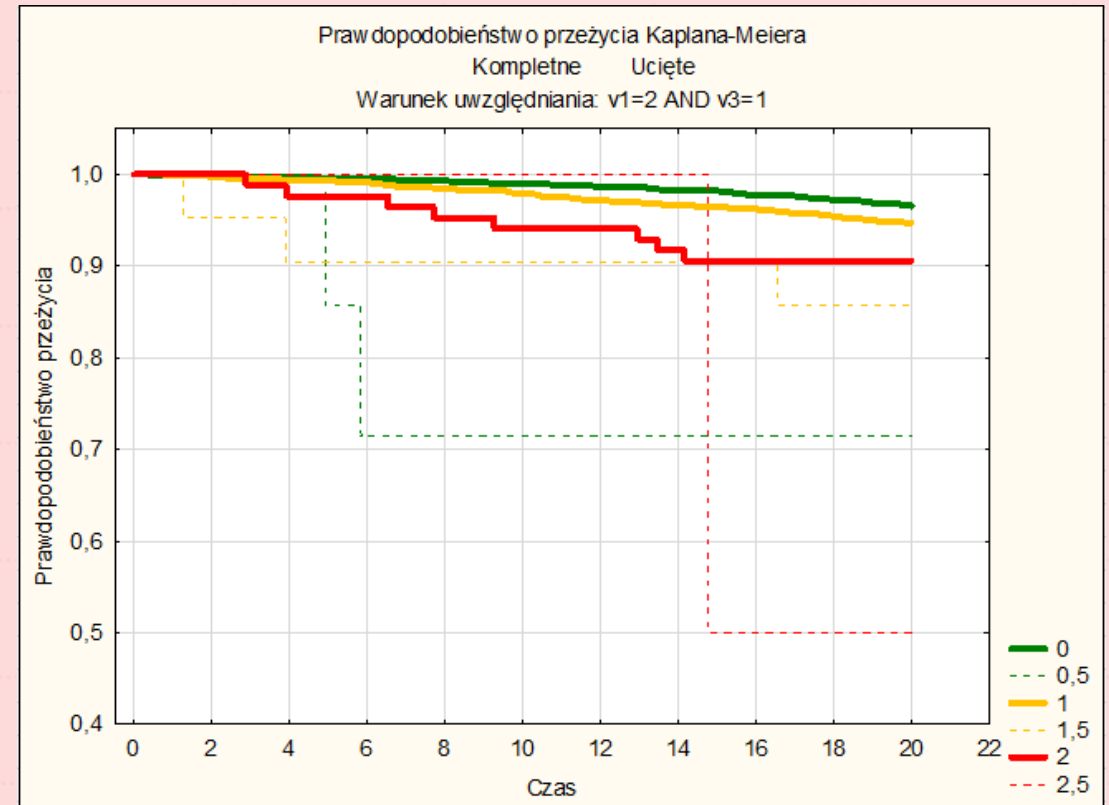


Optimal cholesterol levels (<50 years old)

Male

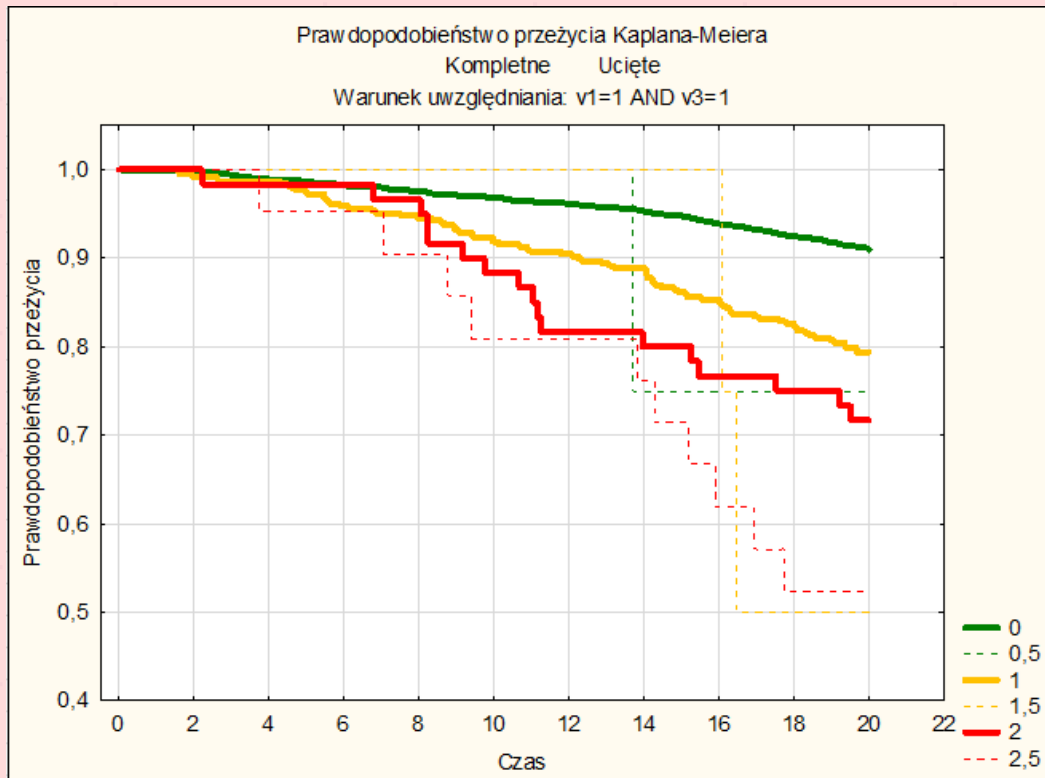


Female

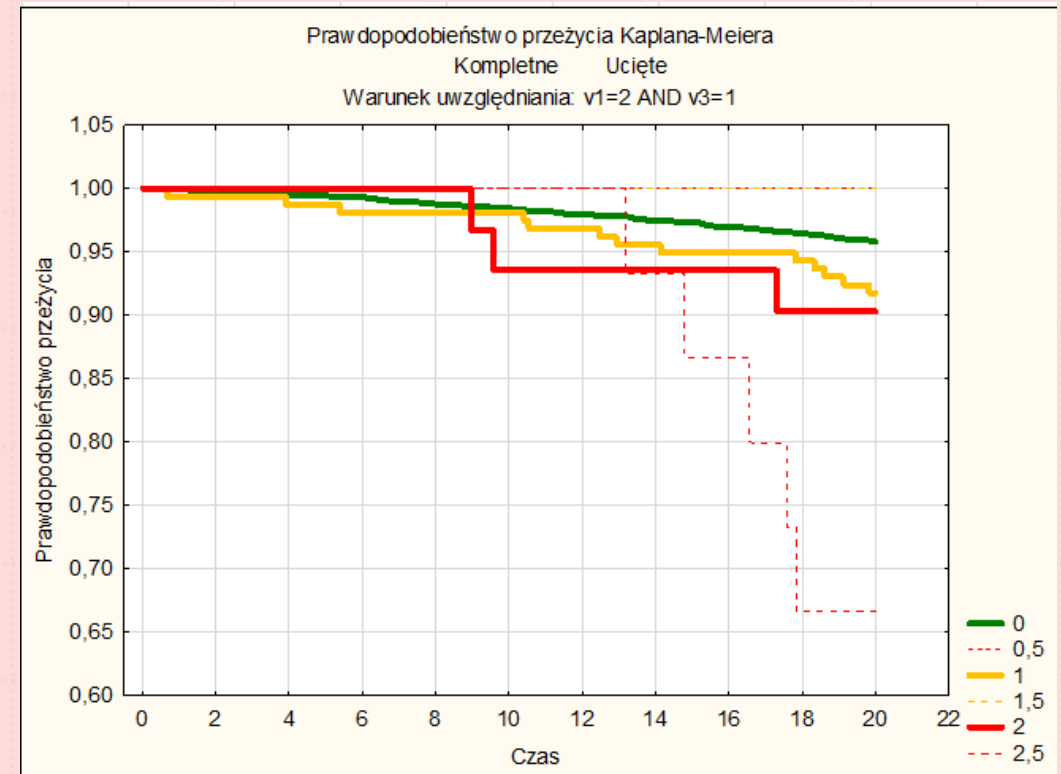


Optimal blood glucose levels (<50 years old)

Male

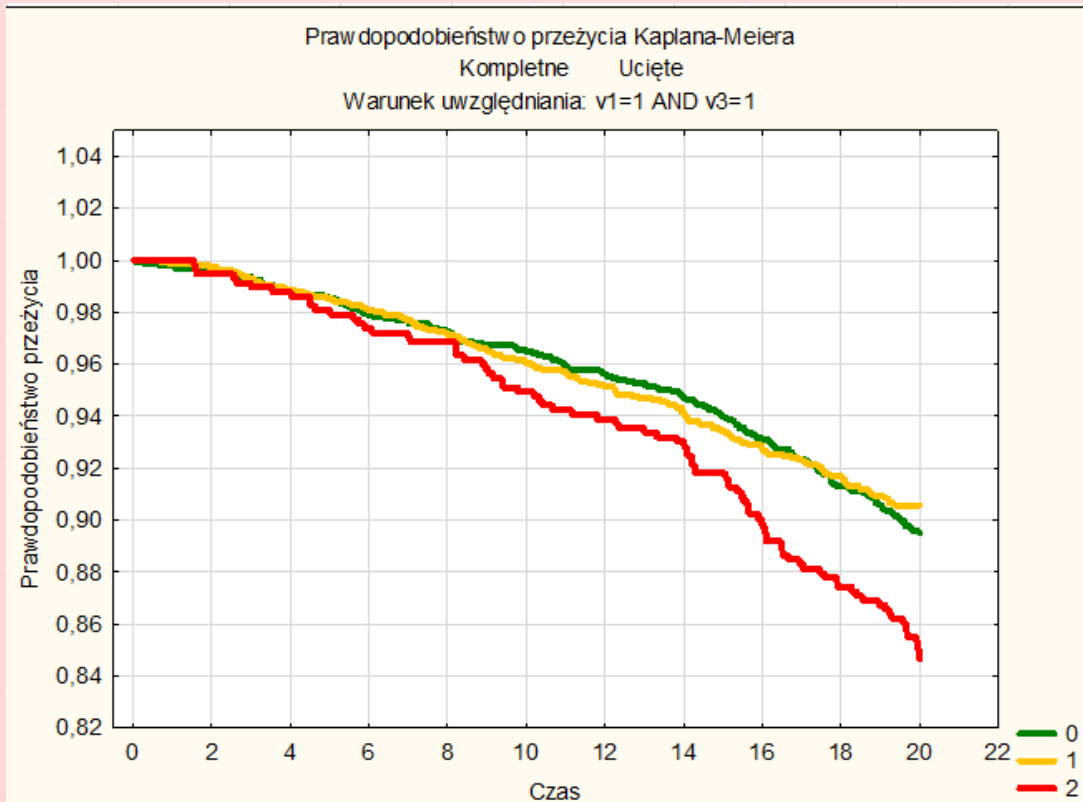


Female

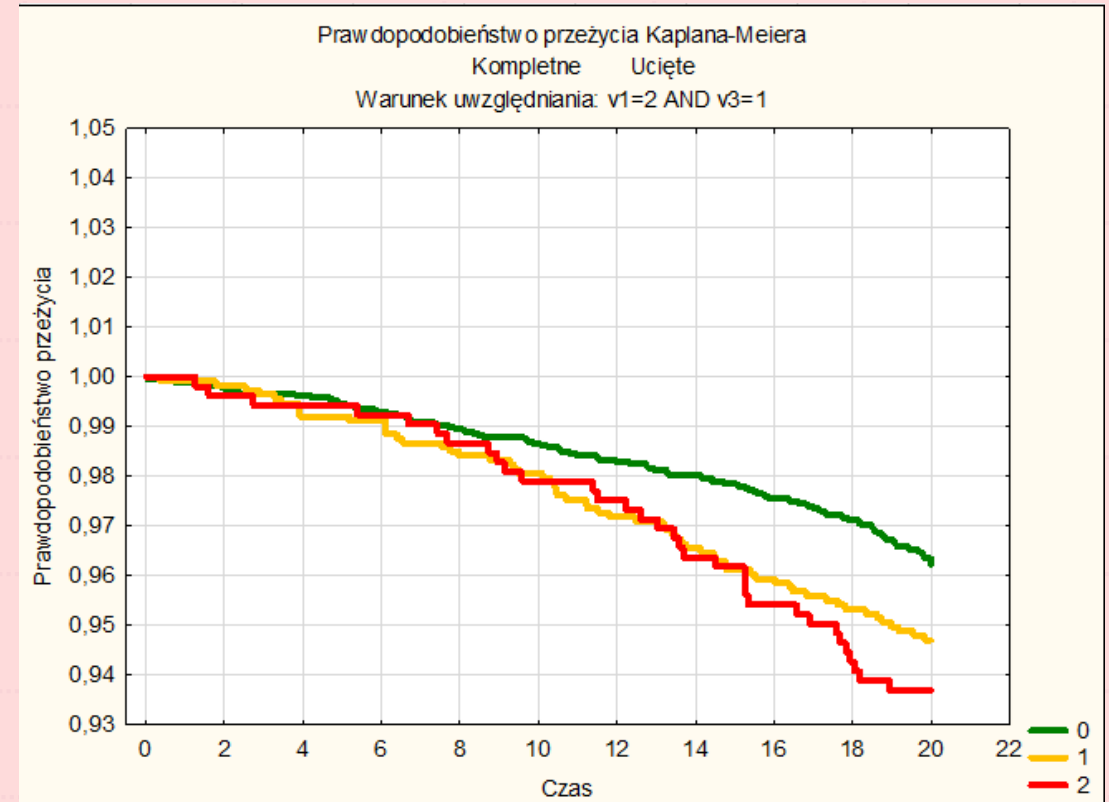


Optimal body weight (<50 years old)

Male

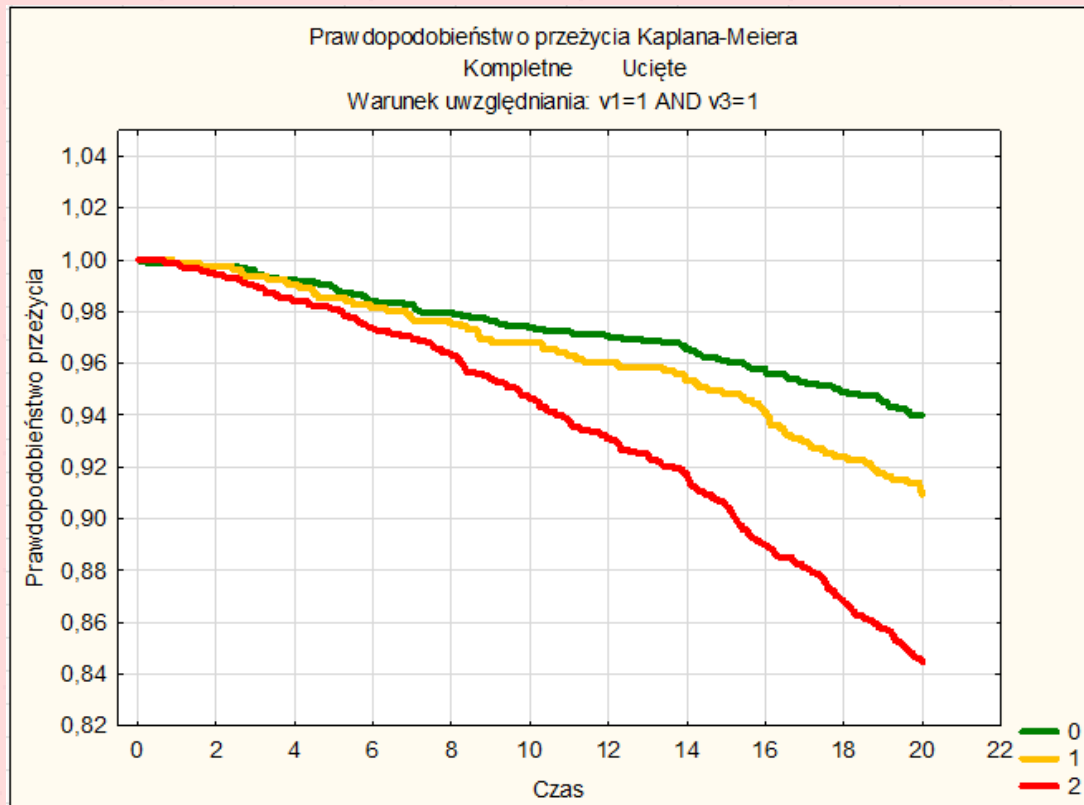


Female

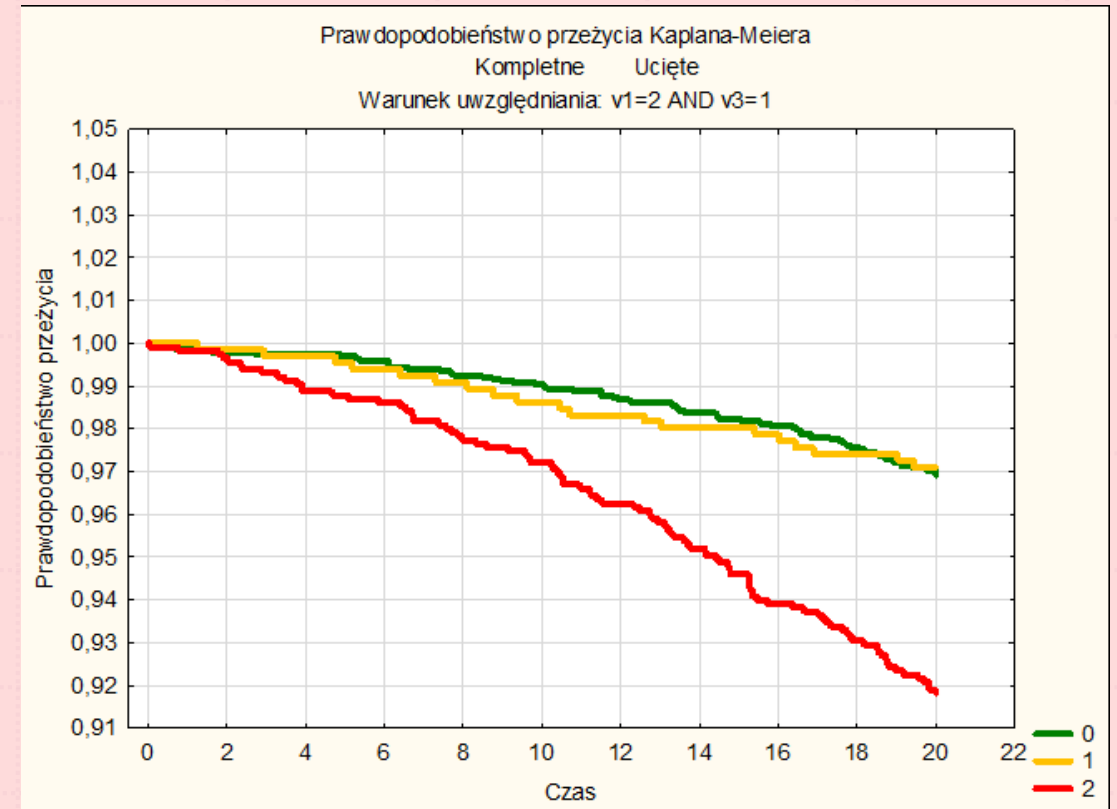


No smoking (<50 years old)

Male

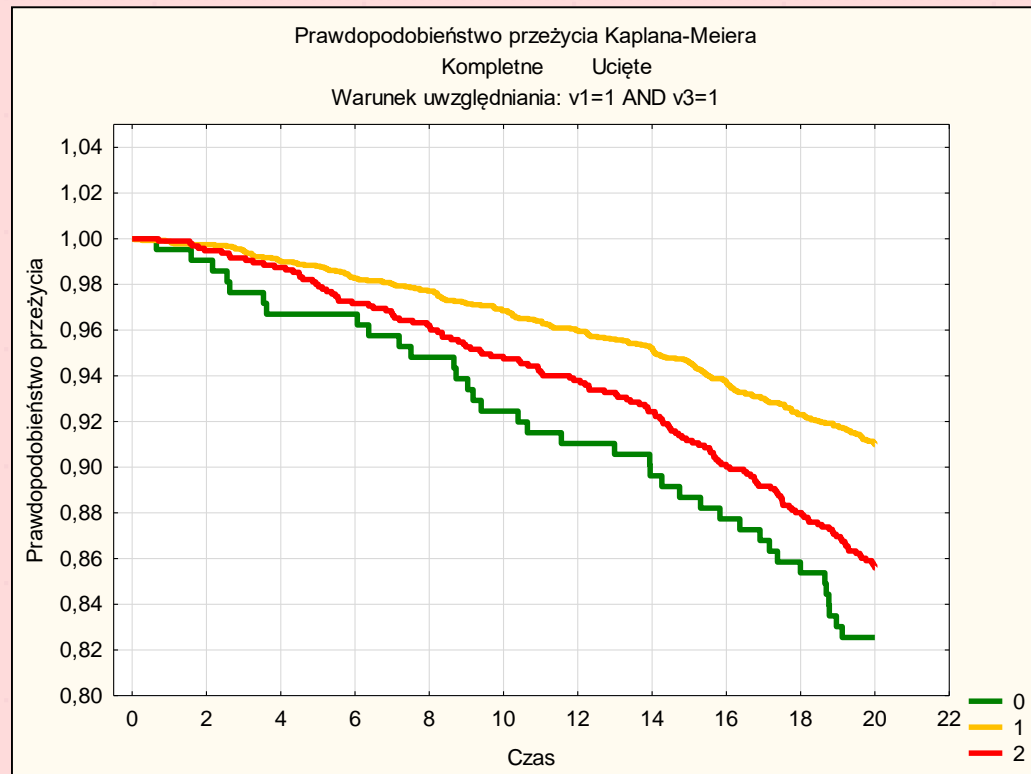


Female

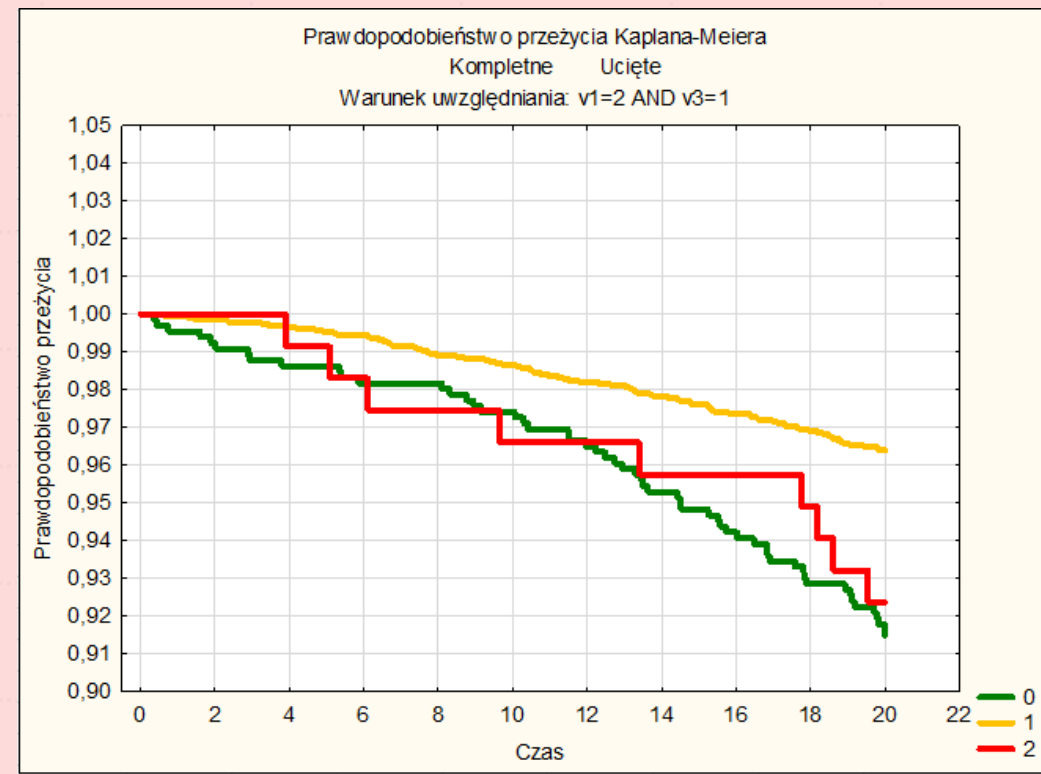


Alcohol consumption (<50 years old)

Male

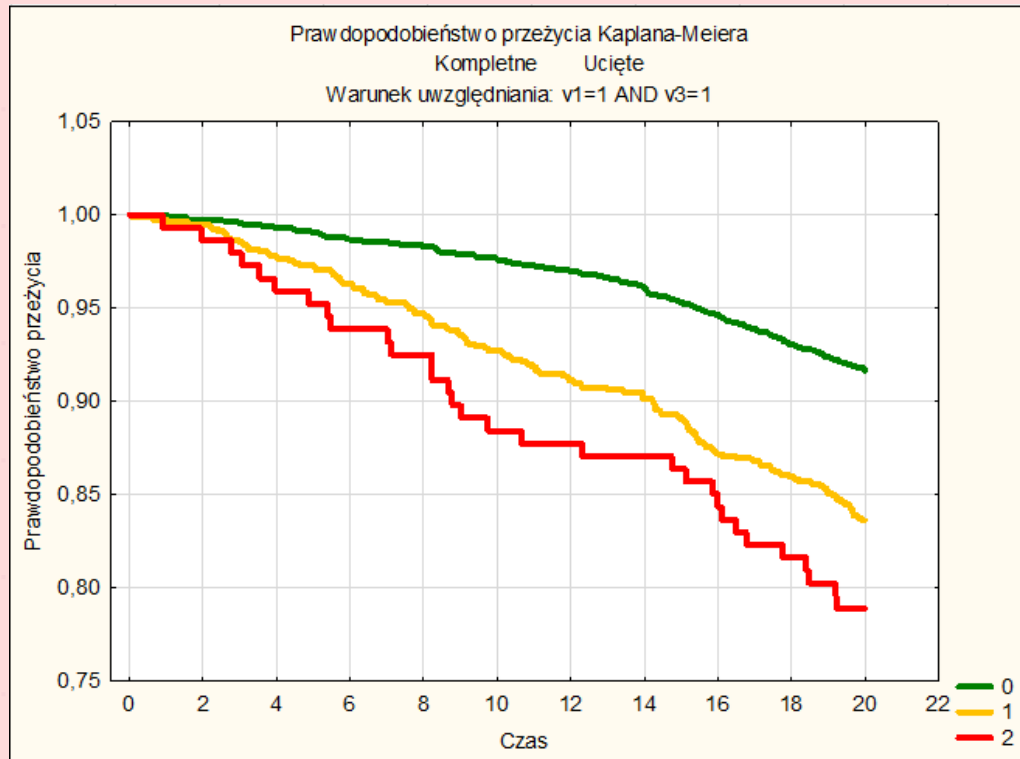


Female

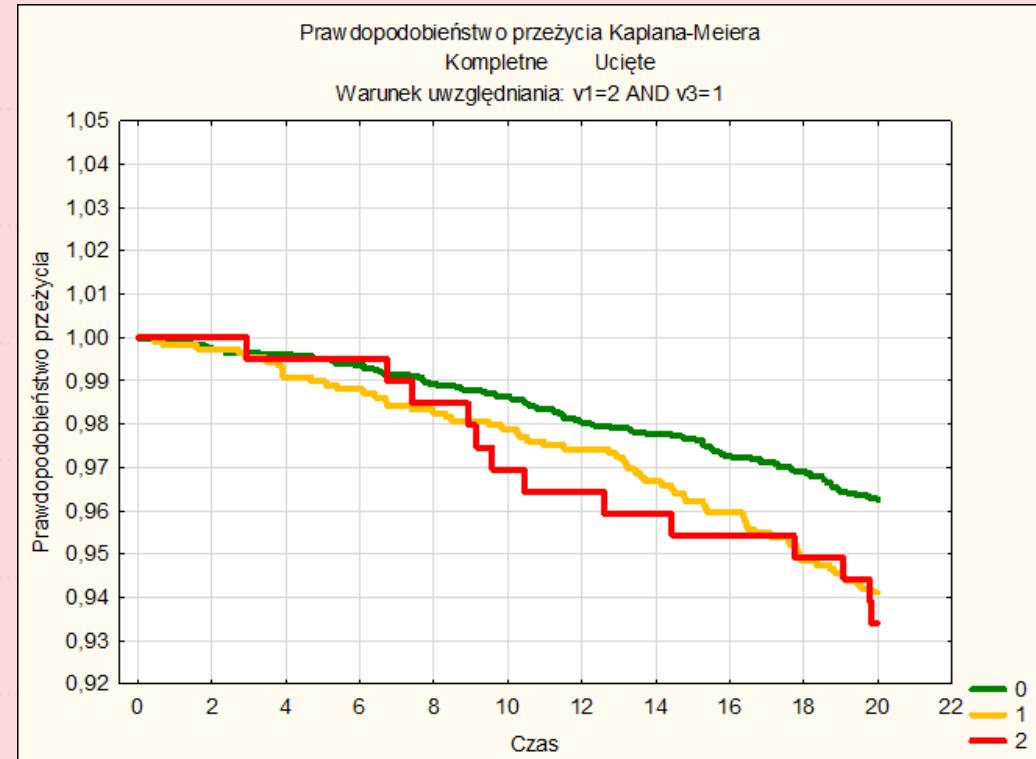


Healthy sleep (<50 years old)

Male

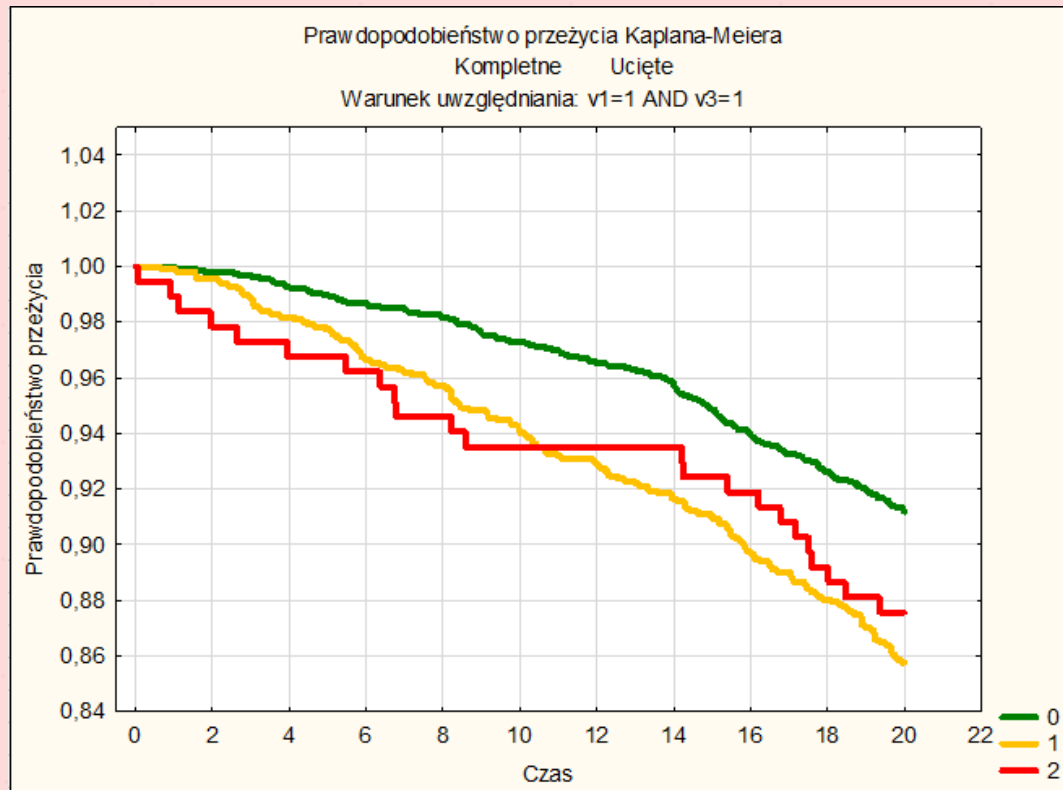


Female

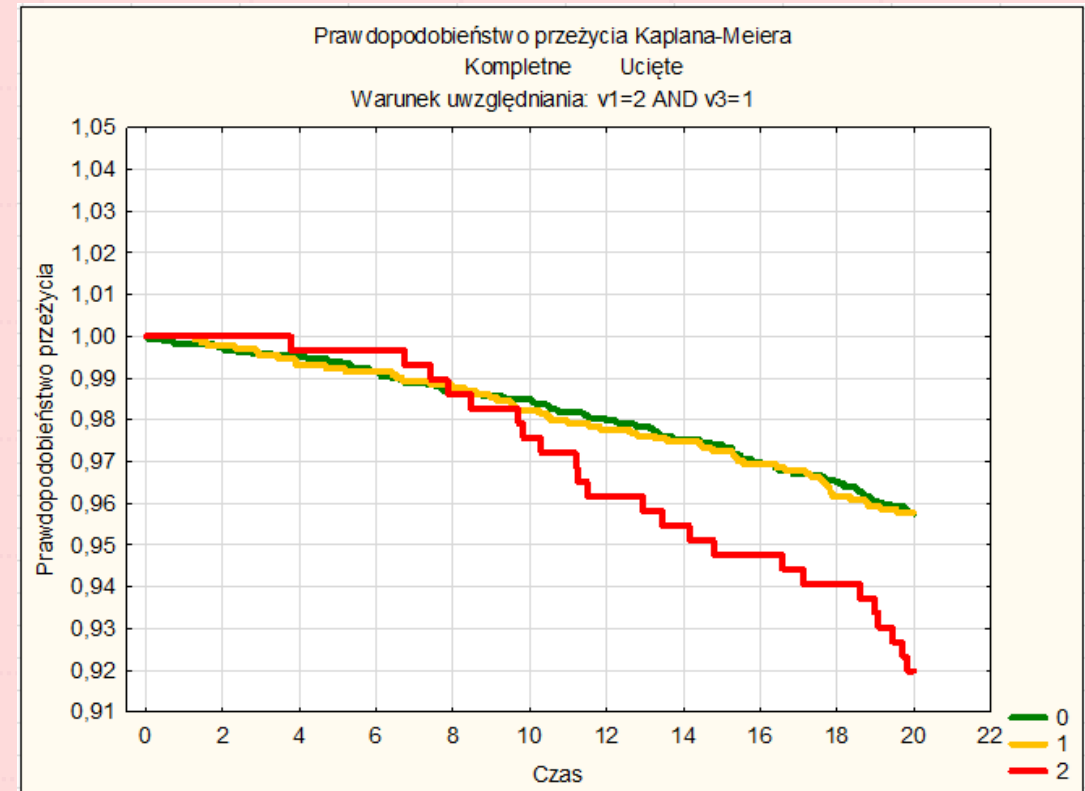


Low stress levels (<50 years old)

Male

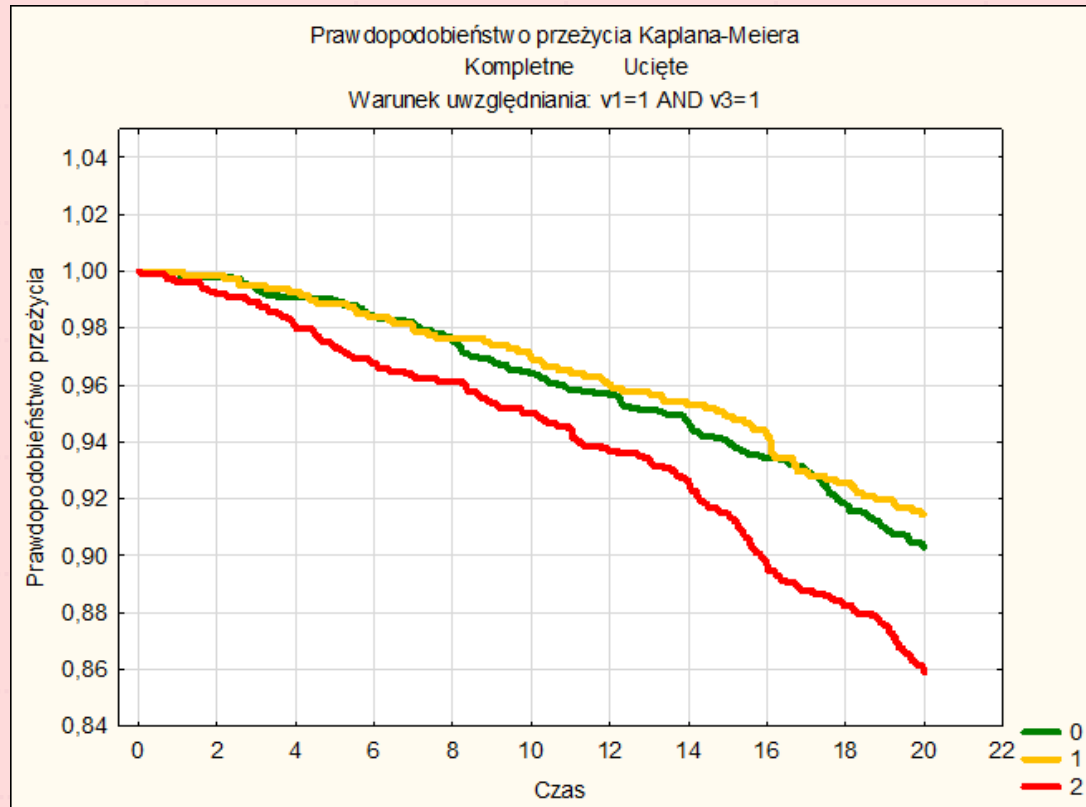


Female

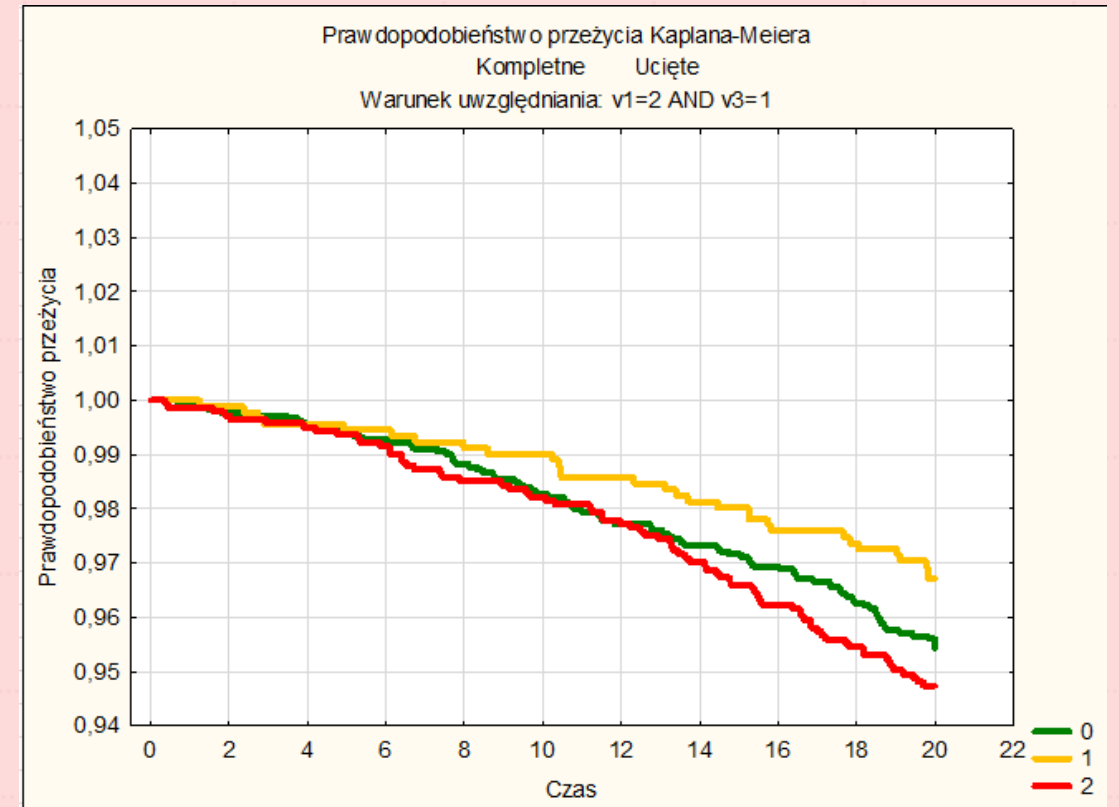


Physical activity (<50 years old)

Male



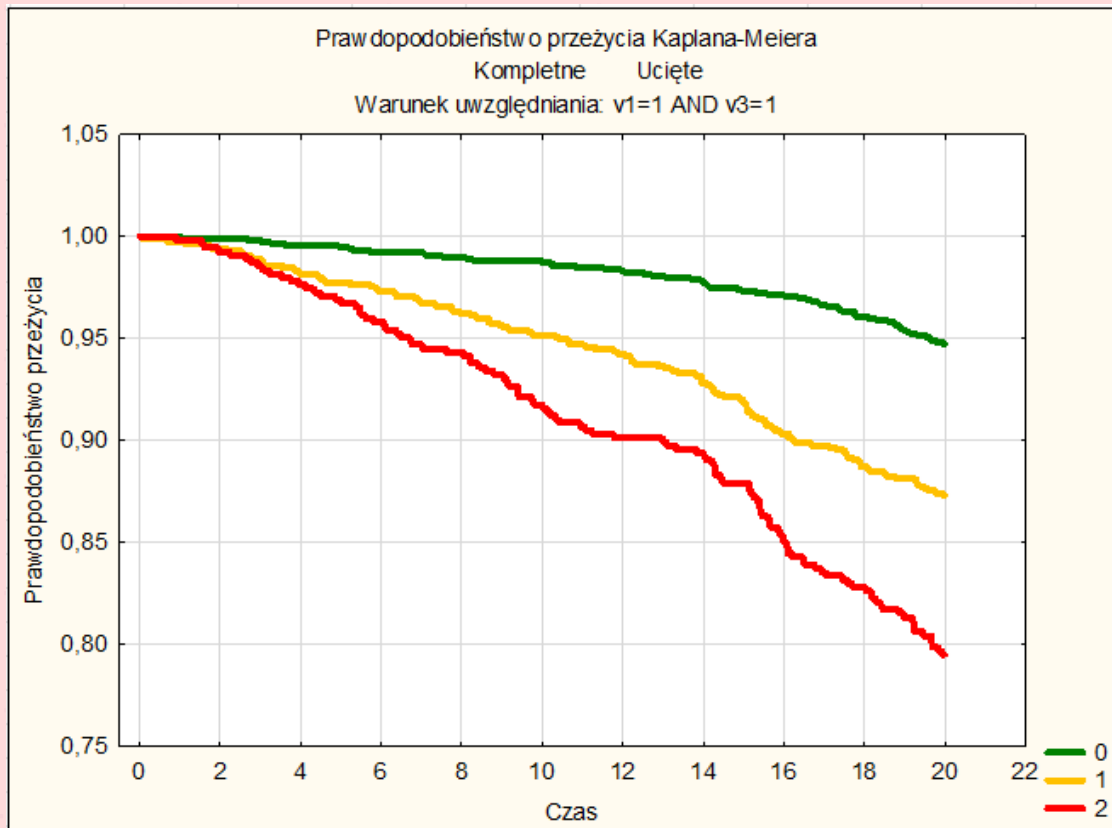
Female



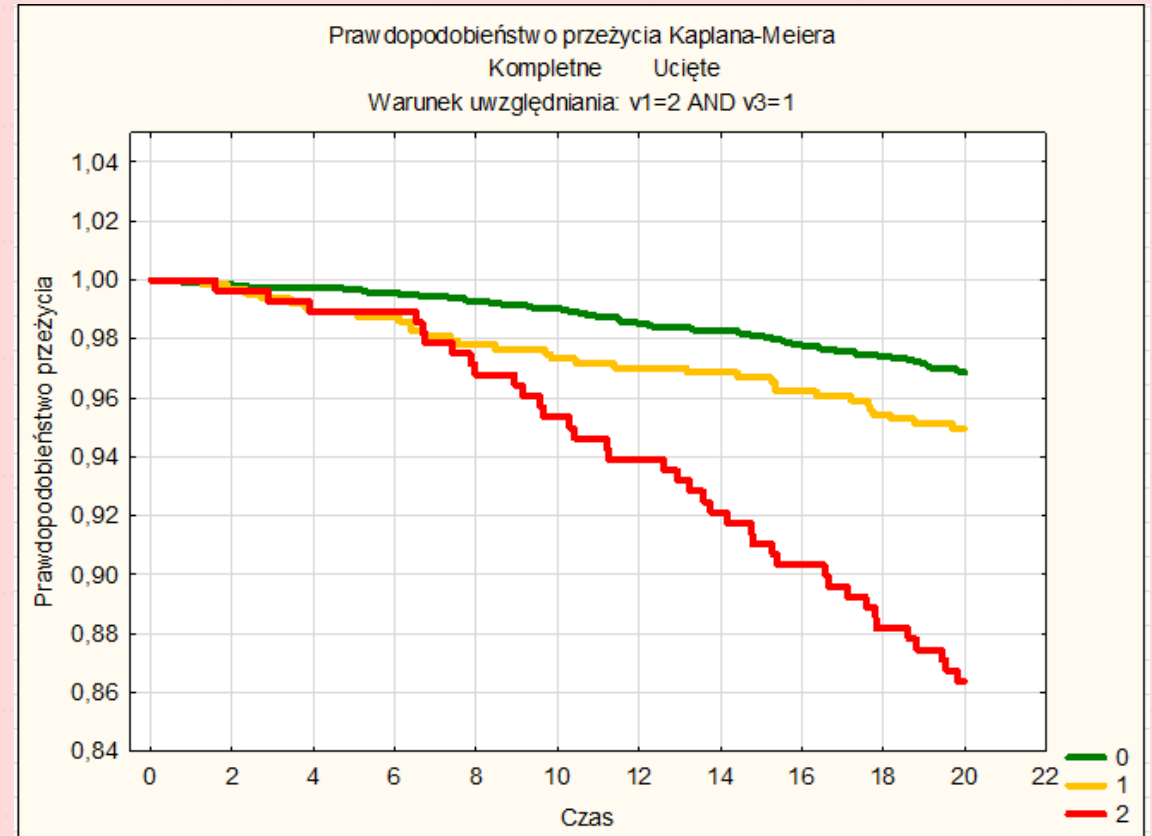
Total points number

0: 0-6 / 1: 7-8 / 2: 9+

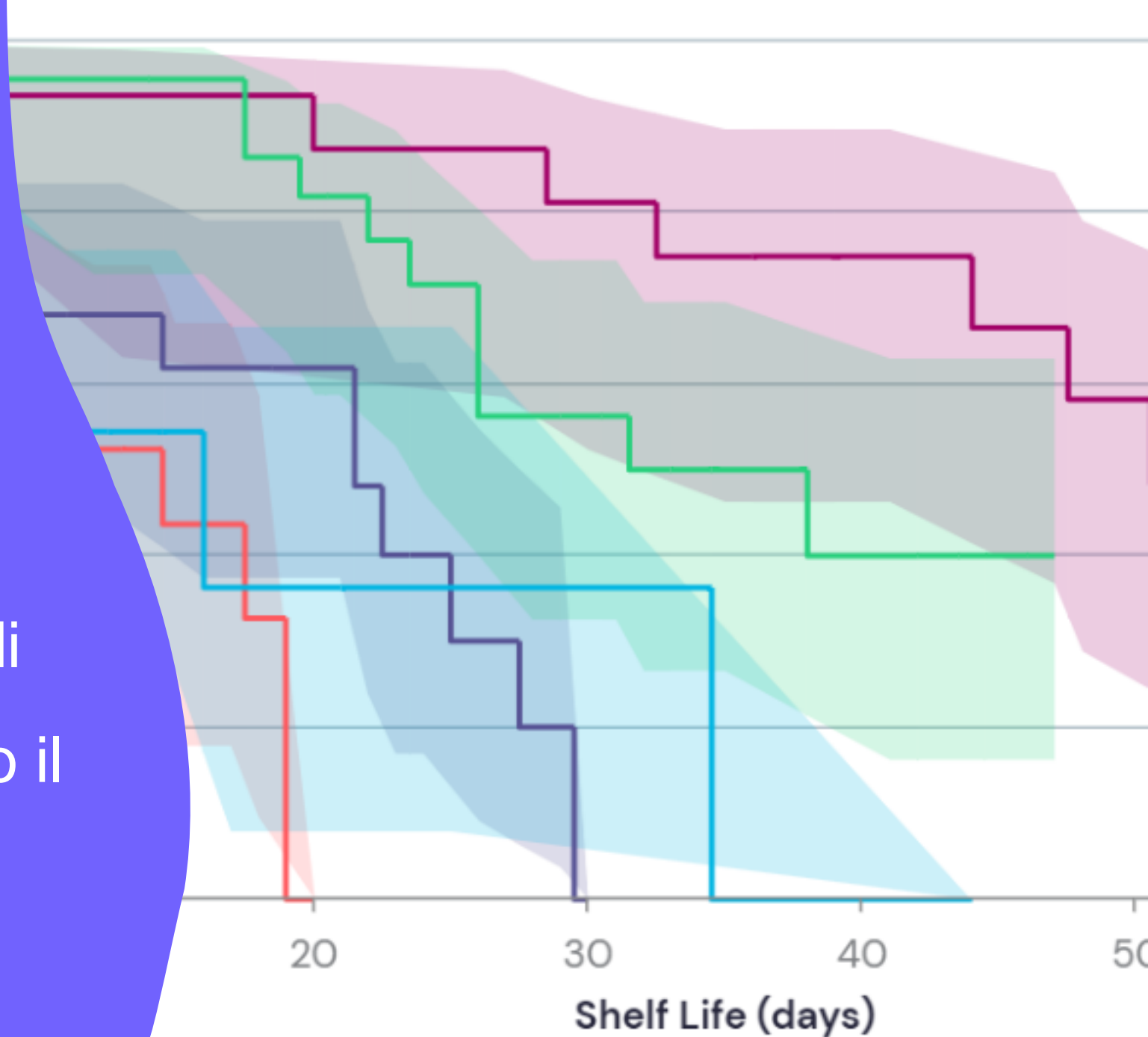
Male



Female



Senza entrare nei
dettagli (troppa teoria
per un solo giorno),
nel modello di Cox
consideriamo il tempo di
sopravvivenza, non solo il
risultato finale:
vivo o morto..



Cox's final model per males

Variables	Beta	SE	t Statistics	Wald	p	HR
Blood pressure	0,677365	0,051290	13,20665	174,4156	0,000000	1,968683
Cholesterol	0,414168	0,086351	4,79632	23,0047	0,000002	1,513111
Glucose	0,662425	0,084820	7,80979	60,9928	0,000000	1,939491
Overweight	0,194666	0,058738	3,31416	10,9837	0,000920	1,214905
Smoking	0,270484	0,050236	5,38429	28,9906	0,000000	1,310598
Physical activity	0,310386	0,050652	6,12787	37,5508	0,000000	1,363952
Sleeping	0,459963	0,078670	5,84671	34,1840	0,000000	1,584016

7 out of 9 analysed variable are statistically significant in multivariable Hazard Cox's model

Cox's final model per females

Variables	Beta	SE	t Statistics	Wald	p	HR
Blood pressure	1,021067	0,065708	15,53953	241,4771	0,000000	2,776155
Cholesterol	0,507336	0,094273	5,38155	28,9611	0,000000	1,660860
Glucose	1,073183	0,098147	10,93442	119,5616	0,000000	2,924673
Overweight	0,303914	0,067597	4,49598	20,2138	0,000007	1,355153
Physical activity	0,498687	0,060829	8,19816	67,2098	0,000000	1,646558
Sleeping	0,392143	0,091891	4,26748	18,2114	0,000020	1,480150

6 out of 9 analysed variable are statistically significant in multivariable Hazard Cox's model

Conclusione

- 1) Possiamo usare la regressione logistica se la variabile dipendente è binaria (0 o 1). Situazione tipica: sano vs. malato / vivo vs. morto / positivo vs. negativo, ecc.
- 2) Nell'analisi di sopravvivenza, il modello di Cox (hazard) è migliore della regressione logistica.
- 3) I fattori di rischio comuni sono statisticamente significativi e legati alla sopravvivenza umana. Dobbiamo pensarci: vogliamo vivere a lungo



Grazie per l'attenzione.
Sarò felice di
rispondere a eventuali
domande, ma **vi prego**
di farle in inglese.

